

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:19:57 ; Search time 16 Seconds  
(without alignments)  
32.266 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193

Perfect score: 62

Sequence: 1 CMPHPRDYDS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 103740

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.rep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.rep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.rep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.rep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.rep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.rep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	45.2	9	3	US-08-159-339A-1107
2	28	45.2	10	3	US-08-159-339A-1077
3	28	45.2	10	3	US-08-159-339A-1108
4	25	40.3	9	3	US-09-139-802-93
5	25	40.3	9	4	US-08-469-260A-567
6	25	40.3	9	4	US-09-659-786-93
7	25	40.3	9	4	US-08-488-446-567
8	25	40.3	9	4	US-08-926-914-93
9	25	40.3	9	4	US-08-467-344A-567
10	25	40.3	8	1	US-08-271-830-46
11	24	38.7	9	2	US-08-483-077C-23
12	24	38.7	9	2	US-08-519-109B-23
13	24	38.7	10	1	US-08-433-783-2
14	24	38.7	10	1	US-08-337-358-2
15	24	38.7	10	5	PCT-US95-12144-2
16	24	38.7	10	5	PCT-US95-07537A-2
17	24	38.7	10	5	PCT-US95-07537-2
18	23	37.1	6	2	US-08-050-235-14
19	23	37.1	6	2	US-08-661-767-14
20	23	37.1	9	4	US-09-125-641-24
21	23	37.1	10	1	US-08-318-970B-8
22	23	37.1	10	1	US-08-481-128-2
23	23	37.1	10	2	US-08-556-597-112
24	23	37.1	10	2	US-09-015-729-2
25	23	37.1	10	4	US-09-551-933-2
26	23	35.5	7	4	US-08-757-425B-72
27	22	35.5	8	5	PCT-US95-11137-25

#### ALIGNMENTS

28	22	35.5	10	2	US-08-633-148-16	Sequence 16, Appl
29	22	35.5	10	3	US-08-836-075A-198	Sequence 198, Appl
30	21	33.9	6	1	US-08-487-006-53	Sequence 53, Appl
31	21	33.9	6	2	US-08-488-659A-53	Sequence 53, Appl
32	21	33.9	6	3	US-08-581-918A-20	Sequence 20, Appl
33	21	33.9	6	3	US-08-346-147B-20	Sequence 20, Appl
34	21	33.9	6	3	US-08-123-934A-19	Sequence 19, Appl
35	21	33.9	6	4	US-08-497-214D-20	Sequence 20, Appl
36	21	33.9	6	4	US-09-874-628-19	Sequence 19, Appl
37	21	33.9	6	5	PCT-US94-10080-19	Sequence 19, Appl
38	21	33.9	7	1	US-07-956-848A-57	Sequence 57, Appl
39	21	33.9	7	1	US-08-471-956-57	Sequence 57, Appl
40	21	33.9	7	2	US-08-437-116A-12	Sequence 12, Appl
41	21	33.9	7	2	US-08-530-792D-12	Sequence 12, Appl
42	21	33.9	8	1	US-08-003-111B-10	Sequence 10, Appl
43	21	33.9	8	1	US-08-271-830-54	Sequence 54, Appl
44	21	33.9	8	1	US-08-014-426-56	Sequence 56, Appl
45	21	33.9	8	1	US-08-261-432-10	Sequence 10, Appl

RESULT 1  
US-08-159-339A-1107  
; Sequence 1107, Application US/08159339A  
; Patent No. 6037135  
GENERAL INFORMATION:  
APPLICANT: Kubo, Ralph T.  
APPLICANT: Grey, Howard W.  
APPLICANT: Settle, Alessandro  
APPLICANT: Celis, Esteban  
TITLE OF INVENTION: HLA Binding peptides and Their  
NUMBER OF SEQUENCES: 1254  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,339A  
FILING DATE: 29-NOV-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/926,666  
FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: US 08/027,746  
FILING DATE: 05-MAR-1993  
APPLICATION NUMBER: US 08/103,396  
FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 018623-005030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
TELEX:  
INFORMATION FOR SEQ ID NO: 1107:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-159-339A-1107

Query Match 45.2%; Score 28; DB 3; Length 9;  
Best Local Similarity 66.7%; Pred. No. 3e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPEHRD 7  
: ||||  
Db 1 LPRHRD 6

RESULT 2

US-08-159-339A-1077  
; Sequence 1077, Application US/08159339A  
; Patent No. 6037135  
; GENERAL INFORMATION:  
; APPLICANT: Kubo, Ralph T.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Celis, Esteban  
; TITLE OF INVENTION: HLA Binding peptides and Their  
; NUMBER OF SEQUENCES: 1254  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,339A  
; FILING DATE: 29-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,666  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: US 08/027,746  
; FILING DATE: 05-MAR-1993  
; APPLICATION NUMBER: US 08/103,396  
; FILING DATE: 06-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 018623-005030US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1077:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-159-339A-1077

Query Match 45.2%; Score 28; DB 3; Length 10;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPEHRD 7  
: ||||  
Db 2 LPRHRD 7

RESULT 3

US-08-159-339A-1108

; Sequence 1108, Application US/08159339A  
; Patent No. 6037135  
; GENERAL INFORMATION:  
; APPLICANT: Kubo, Ralph T.  
; APPLICANT: Sette, Alessandro  
; APPLICANT: Celis, Esteban  
; TITLE OF INVENTION: HLA Binding peptides and Their  
; NUMBER OF SEQUENCES: 1254  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,339A  
; FILING DATE: 29-NOV-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,666  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: US 08/027,746  
; FILING DATE: 05-MAR-1993  
; APPLICATION NUMBER: US 08/103,396  
; FILING DATE: 06-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 018623-005030US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1108:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-159-339A-1108

Query Match 45.2%; Score 28; DB 3; Length 10;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPEHRD 7  
: ||||  
Db 1 LPRHRD 6

RESULT 4

US-09-139-802-93  
; Sequence 93, Application US/09139802  
; Patent No. 6180084  
; GENERAL INFORMATION:  
; APPLICANT: Ruoslahti, Erkki  
; APPLICANT: Pasqualini, Renata  
; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
; FILE REFERENCE: P-LJ 3203  
; CURRENT APPLICATION NUMBER: US/09/139,802  
; CURRENT FILING DATE: 1998-08-25

EARLIER APPLICATION NUMBER: 08/526,914  
EARLIER FILING DATE: 1997-09-10  
EARLIER APPLICATION NUMBER: 08/710,067  
EARLIER FILING DATE: 1996-09-10  
NUMBER OF SEQ ID NOS: 226  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 93  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-139-802-93

Query Match  
Best Local Similarity 40.3%; Score 25; DB 3; Length 9;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PEHR 6  
Db 2 PEHR 5

RESULT 5  
US-08-469-260A-567  
Sequence 567, Application US/08469260A  
Patent No. 6451578  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMU J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUERHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSER: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 567:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-469-260A-567

Query Match  
Best Local Similarity 40.3%; Score 25; DB 4; Length 9;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 EHRDYD 9  
Db 2 QHRDRD 7

RESULT 6  
US-09-659-786-93  
Sequence 93, Application US/09659786  
Patent No. 6491894  
GENERAL INFORMATION:  
APPLICANT: Ruoslahti, Erkki  
APPLICANT: Pasqualini, Renata  
TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
FILE REFERENCE: P-1J 3203  
CURRENT APPLICATION NUMBER: US/09/659,786  
CURRENT FILING DATE: 2000-09-11  
PRIOR APPLICATION NUMBER: 08/526,914  
PRIOR FILING DATE: 1997-09-10  
PRIOR APPLICATION NUMBER: 08/710,067  
PRIOR FILING DATE: 1996-09-10  
NUMBER OF SEQ ID NOS: 226  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 93  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-659-786-93

Query Match  
Best Local Similarity 40.3%; Score 25; DB 4; Length 9;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PEHR 6  
Db 2 PEHR 5

RESULT 7  
US-08-488-446-567  
Sequence 567, Application US/08488446  
Patent No. 6558888  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMU J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUERHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSER: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA

ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,446  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: PORZEMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 567:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-486-446-567

Query Match 40.3%; Score 25; DB 4; Length 9;  
Best Local Similarity 66.7%; Pred. No. 3e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 EHRDYP 9  
Db 2 QHRDRD 7

RESULT 8  
US-08-926-914-93  
Sequence 93, Application US/08926914  
Patent No. 6576239  
GENERAL INFORMATION:  
APPLICANT: Ruoslahti, Erkki  
APPLICANT: Pasqualini, Renata  
TITLE OF INVENTION: Tumor Homing Molecules, Conjugates  
TITLE OF INVENTION: Derived therefrom, and Methods of Using Same  
NUMBER OF SEQUENCES: 199  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/926,914  
FILING DATE: 10-SEP-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2725  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids

TYPE: amino acid  
TOPOLOGY: both  
MOLECULE TYPE: peptide  
US-08-926-914-93

Query Match 40.3%; Score 25; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PEHR 6  
Db 2 PEHR 5

RESULT 9  
US-08-467-344A-567  
Sequence 567, Application US/08467344A  
Patent No. 6586568  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
TAMI J. PILOT-MATIAS  
GEORGE J. DAWSON  
GEORGE G. SCHLAUDER  
SURESH M. DESAI  
THOMAS P. LEARY  
ANTHONY SCOTT MUEHRHOFF  
JAMES C. ERKER  
SHERI L. BUIJK  
ISA K. MUSHAMMAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,344A  
FILING DATE: 07-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/424,550  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: PORZEMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 567:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 567:  
US-08-467-344A-567

Query Match 40.3%; Score 25; DB 4; Length 9;  
Best Local Similarity 66.7%; Pred. No. 3e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 EHRDYP 9  
Db 2 QHRDRD 7



Db 2 QHRDRD 7

RESULT 10

US-08-271-830-46

Sequence 46, Application US/08271830

Patent No. 5510332

GENERAL INFORMATION:

APPLICANT: Kogan, Timothy P.

APPLICANT: Ren, Kaijun

APPLICANT: Vanderslice, Peter

APPLICANT: Beck, Pamela J.

TITLE OF INVENTION: A PROCESS OF INHIBITING THE BINDING OF THE

TITLE OF INVENTION: INTEGRIN '4.1 TO VCAM OR FIBRONECTIN AND

TITLE OF INVENTION: LINEAR PEPTIDES THEREFOR

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dressler, Goldsmith, Shore & Milnamow, Ltd.

STREET: 180 No. 5510332th Steetson, Suite 4700

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/271,830

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5510332thrup, Thomas E.

REGISTRATION NUMBER: 33,268

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 616-5400

TELEFAX: (312) 616-5460

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 8 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 8

OTHER INFORMATION: /label= Xaa

OTHER INFORMATION: /note= "Xaa=Pro-NH2."

US-08-271-830-46

Query Match 38.7%; Score 24; DB 1; Length 8;

Best Local Similarity 80.0%; Pred. No. 3e+05; 1; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PHRD 7

Db 2 PHRD 6

RESULT 11

US-08-483-077C-23

Sequence 23, Application US/08483077C

Patent No. 5811391

GENERAL INFORMATION:

APPLICANT: Arthenius, Thomas S.

APPLICANT: Tempczyk, Anna

APPLICANT: Elices, Mariano J.

APPLICANT: Zheng, Zhong-Li

TITLE OF INVENTION: Cyclic CS-1 Peptidomimetics,

TITLE OF INVENTION: Compositions and Methods of Using Same

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,077C

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-CY 1647

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

TOPOLOGY: circular

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 9

OTHER INFORMATION: /note= "The hydroxy group attached

OTHER INFORMATION: to the carbonyl group is replaced by an amine

OTHER INFORMATION: group."

US-08-483-077C-23

Query Match 38.7%; Score 24; DB 2; Length 9;

Best Local Similarity 50.0%; Pred. No. 3e+05; 3; Indels 0; Gaps 0;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CPEHRD 8

Db 1 CPEHRD 8

RESULT 12

US-08-519-109B-23

Sequence 23, Application US/08519109B

Patent No. 5869448

GENERAL INFORMATION:

APPLICANT: Arthenius, Thomas S.

APPLICANT: Tempczyk, Anna

APPLICANT: Elices, Mariano J.

APPLICANT: Zheng, Zhong-Li

TITLE OF INVENTION: Cyclic CS-1 Peptidomimetics,

TITLE OF INVENTION: Compositions and Methods of Using Same

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell & Flores LLP

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: United States

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/519,109B

FILING DATE: 25-AUG-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CY 1795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: circular  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 9  
OTHER INFORMATION: /note="The hydroxy group attached  
OTHER INFORMATION: to the carbonyl group is replaced by an amine  
OTHER INFORMATION: group."  
US-08-519-109B-23

Query Match 38.7%; Score 24; DB 2; Length 9;  
Best Local Similarity 50.0%; Pred. No. 3e+05;  
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CMPEHRDY 8  
DB 1 CGPFPDF 8

RESULT 13  
US-08-433-783-2  
Sequence 2, Application US/08433783  
Patent No. 5770431  
GENERAL INFORMATION:  
APPLICANT: Liu, Chi-Li  
APPLICANT: Adams, Lee F.  
APPLICANT: Lufburrow, Patricia A.  
APPLICANT: Thomas, Michael D.  
TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS STRAINS  
TITLE OF INVENTION: ACTIVE AGAINST LEPIDOPTERAN AND COLEOPTERAN PESTS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5770431, disk of No. 5770431th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/433,783  
FILING DATE: 03-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,358  
FILING DATE: 10-NOV-1994  
APPLICATION NUMBER: US 08/264,100  
FILING DATE: 22-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,651  
FILING DATE: 09-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/166,391  
FILING DATE: 13-DEC-1993  
APPLICATION NUMBER: US 07/991,073  
FILING DATE: 15-DEC-1992  
ATTORNEY/AGENT INFORMATION:

NAME: Agiris Dr., Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3778.230-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-433-783-2

Query Match 38.7%; Score 24; DB 1; Length 10;  
Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPEHRDYD 9  
DB 1 MGHKKNFD 8

RESULT 14  
US-08-337-358-2  
Sequence 2, Application US/08337358  
Patent No. 5879676  
GENERAL INFORMATION:  
APPLICANT: Liu, Chi-Li  
APPLICANT: Adams, Lee F.  
APPLICANT: Lufburrow, Patricia A.  
APPLICANT: Thomas, Michael D.  
TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS STRAINS  
TITLE OF INVENTION: ACTIVE AGAINST LEPIDOPTERAN AND COLEOPTERAN PESTS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5879676, disk of No. 5879676th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,358  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/264,100  
FILING DATE: 22-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,651  
FILING DATE: 09-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/166,391  
FILING DATE: 13-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,073  
FILING DATE: 15-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agiris Dr., Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3778.230-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-337-358-2

Query Match 38.7%; Score 24; DB 2; Length 10;  
Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPEHRDYL 9  
| | | | |  
| | | | |  
Db 1 MGHKKNFD 8

## RESULT 15

PCT-US93-12144-2  
Sequence 2, Application PC/TUS9312144  
GENERAL INFORMATION:  
APPLICANT: Liu, Chi-Li  
APPLICANT: Adams, Lee F.  
APPLICANT: Luburrow, Patricia A.  
APPLICANT: Thomas, Michael D.  
TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS STRAINS  
TITLE OF INVENTION: ACTIVE AGAINST LEPIDOPTERAN AND COLEOPTERAN PESTS  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Novo Nordisk of North America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/12144  
FILING DATE: 13-DEC-1993  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,073  
FILING DATE: 15-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agria, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3778.204-WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-12144-2

Query Match 38.7%; Score 24; DB 5; Length 10;  
Best Local Similarity 37.5%; Pred. No. 2.3e+02;  
Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPEHRDYL 9  
| | | | |  
| | | | |  
Db 1 MGHKKNFD 8

Search completed: September 9, 2004, 06:22:33  
Job time: 18 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:48:37 ; Search time 32 Seconds  
(without alignments)  
120,998 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75  
Perfect score: 373  
Sequence: 1 MAEIGEDLDKSDVSLIFLM.....LEKCIKNIRHLDKTKIKQY 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 252506

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUTS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	26.5	75	1	US-08-665-220-66 Sequence 66, App1
2	91	24.4	68	4	US-09-180-167A-13 Sequence 13, App1
3	91	24.4	68	4	US-09-033-524B-13 Sequence 13, App1
4	90	24.1	75	3	US-09-291-682-66 Sequence 66, App1
5	80	21.4	66	4	US-09-180-167A-12 Sequence 12, App1
6	80	21.4	66	4	US-09-033-524B-12 Sequence 12, App1
7	69	18.5	64	4	US-09-180-167A-31 Sequence 31, App1
8	69	18.5	64	4	US-09-033-524B-31 Sequence 31, App1
9	66	17.7	67	4	US-09-180-167A-10 Sequence 10, App1
10	66	17.7	67	4	US-09-033-524B-10 Sequence 10, App1
11	51.5	13.8	47	4	US-09-823-266-15 Sequence 15, App1
12	51	13.7	47	4	US-09-823-266-13 Sequence 13, App1
13	51	13.7	67	4	US-09-180-167A-11 Sequence 11, App1
14	51	13.7	67	4	US-09-033-524B-11 Sequence 11, App1
15	50.5	13.1	69	4	US-09-621-976-7353 Sequence 7353, App1
16	49	13.1	54	4	US-09-621-976-4976 Sequence 4976, App1
17	49	13.1	68	4	US-08-974-549A-181 Sequence 181, App1
18	49	13.1	72	4	US-09-621-976-7354 Sequence 7354, App1
19	48.5	13.0	70	3	US-09-367-953B-107 Sequence 107, App1
20	48	12.9	67	2	US-08-142-551B-9 Sequence 9, App1
21	47.5	12.7	47	4	US-09-823-266-14 Sequence 14, App1
22	47	12.6	66	2	US-08-867-087B-40 Sequence 40, App1
23	46	12.3	41	3	US-08-974-549A-181 Sequence 181, App1
24	46	12.3	41	4	US-09-403-181B-181 Sequence 181, App1
25	46	12.3	41	4	US-09-721-456-181 Sequence 181, App1
26	46	12.3	49	3	US-09-382-155-4 Sequence 4, App1
27	46	12.3	49	3	US-09-074-044A-4 Sequence 4, App1

28	45.5	12.2	73	4	US-09-125-635-3 Sequence 3, App1
29	44.5	11.9	66	4	US-09-732-210-829 Sequence 829, App1
30	44.5	11.9	70	4	US-09-543-681A-7547 Sequence 7547, App1
31	44.5	11.9	75	4	US-08-858-207A-460 Sequence 460, App1
32	44	11.8	35	3	US-09-082-279B-276 Sequence 276, App1
33	44	11.8	35	3	US-09-082-279B-276 Sequence 276, App1
34	44	11.8	35	4	US-09-315-304B-276 Sequence 276, App1
35	44	11.8	35	4	US-09-315-304B-276 Sequence 276, App1
36	44	11.8	35	4	US-09-834-784-276 Sequence 276, App1
37	44	11.8	35	4	US-09-834-784-276 Sequence 276, App1
38	44	11.8	35	4	US-09-515-965A-276 Sequence 276, App1
39	44	11.8	35	4	US-09-515-965A-276 Sequence 276, App1
40	44	11.8	35	4	US-09-350-641C-276 Sequence 276, App1
41	44	11.8	35	4	US-09-350-641C-276 Sequence 276, App1
42	44	11.8	39	2	US-08-247-809A-8 Sequence 8, App1
43	44	11.8	39	2	US-08-711-728-8 Sequence 8, App1
44	43.5	11.7	71	1	US-08-280-443-18 Sequence 18, App1
45	43.5	11.7	71	1	US-08-457-459-18 Sequence 18, App1

ALIGNMENTS

RESULT 1  
US-08-665-220-66  
Sequence 66, Application US/08665220  
Patent No. 5786173  
GENERAL INFORMATION:  
APPLICANT: Alimemti, Emad S.  
APPLICANT: Fernandes-Alimemti, Teresa  
APPLICANT: Litwack, Gerald  
APPLICANT: Armstrong, Robert  
APPLICANT: Tomaselli, Kevin  
TITLE OF INVENTION: Mch4 and Mch5, Apoptotic Proteases,  
TITLE OF INVENTION: Nucleic Acids Encoding and Methods of Use  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,220  
FILING DATE: 14-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/616,408  
FILING DATE: 19-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-ID 2165  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..75  
OTHER INFORMATION: /note= "Mch5 A"  
US-08-665-220-66

Thu Sep 9 07:45:53 2004

us-10-713-208-6\_copy\_1\_75.ra1

Page 2

Query Match 26.5%; Score 91; DB 1; Length 75;  
Best Local Similarity 36.4%; Pred. No. 4.5e-05;  
Matches 24; Conservative 13; Mismatches 23; Indels 0; Gaps 0;

Qy 3 EIGEDLDKSDVSSLIIFLMKDYMGKISKEKSFIDLVELEKLNVAPODLDLEKLN 62  
Db 6 DIEEQDSEDLASTKFLSLDTIPQRKQEPITKALMLFORLOEKWLEESNLSFLKELLFR 65

Qy 63 IRRIDL 68  
Db 66 INRLDL 71

RESULT 2  
US-09-180-167A-13  
Sequence 13, Application US/09180167A

Patent No. 6558950  
GENERAL INFORMATION:  
APPLICANT: Gordon C. Shore et al.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
FILE REFERENCE: 50013/004003  
CURRENT APPLICATION NUMBER: US/09/180,167A  
CURRENT FILING DATE: 1999-09-10  
PRIOR APPLICATION NUMBER: CA 2,198,988  
PRIOR FILING DATE: 1997-03-03  
PRIOR APPLICATION NUMBER: PCT/IB98/00706  
PRIOR FILING DATE: 1998-03-02  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-180-167A-13

Query Match 24.4%; Score 91; DB 4; Length 68;  
Best Local Similarity 26.8%; Pred. No. 0.00039;  
Matches 15; Conservative 23; Mismatches 18; Indels 0; Gaps 0;

Qy 3 EIGEDLDKSDVSSLIIFLMKDYMGKISKEKSFIDLVELEKLNVAPODLDLEK 58  
Db 11 QISEVSRSEIRSEKFLQGEISKCLDDDMNLDFIEMERKVLIGSEKDLIKR 66

RESULT 3  
US-09-033-524B-13  
Sequence 13, Application US/09033524B

Patent No. 6607880  
GENERAL INFORMATION:  
APPLICANT: Gordon C. Shore et al.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
FILE REFERENCE: 50013/004002  
CURRENT APPLICATION NUMBER: US/09/033,524B  
CURRENT FILING DATE: 1998-03-02  
PRIOR APPLICATION NUMBER: CA 2,198,988  
PRIOR FILING DATE: 1997-03-03  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-033-524B-13

Query Match 24.4%; Score 91; DB 4; Length 68;  
Best Local Similarity 26.8%; Pred. No. 0.00039;  
Matches 15; Conservative 23; Mismatches 18; Indels 0; Gaps 0;

Qy 3 EIGEDLDKSDVSSLIIFLMKDYMGKISKEKSFIDLVELEKLNVAPODLDLEK 58  
Db 11 QISEVSRSEIRSEKFLQGEISKCLDDDMNLDFIEMERKVLIGSEKDLIKR 66

RESULT 4  
US-09-291-692-66  
Sequence 66, Application US/09291692

Patent No. 6287795  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
APPLICANT: Fernandes-Alnemri, Teresa  
APPLICANT: Litwack, Gerald  
APPLICANT: Armstrong, Robert  
APPLICANT: Tomaselli, Kevin  
TITLE OF INVENTION: MCH4 AND MCH5, APOPTOTIC PROTEASE,  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: Use  
ZIP: 98104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/291,692  
FILING DATE: 04-13-1999  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Christensen, William T.  
REGISTRATION NUMBER: 44,614  
REFERENCE/DOCKET NUMBER: 480140.424C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..75  
OTHER INFORMATION: /note= "Mch5 A"

US-09-291-692-66

Query Match 24.1%; Score 90; DB 3; Length 75;  
Best Local Similarity 34.8%; Pred. No. 0.00059;  
Matches 23; Conservative 13; Mismatches 30; Indels 0; Gaps 0;

Qy 3 EIGEDLDKSDVSSLIIFLMKDYMGKISKEKSFIDLVELEKLNVAPODLDLEKLN 62  
Db 6 DIEEQDSEDLASTKFLSLDTIPQRKQEPITKALMLFORLOEKWLEESNLSFLKELLFR 65

Qy 63 IRRIDL 68  
Db 66 INRLDL 71

RESULT 5  
US-09-180-167A-12  
Sequence 12, Application US/09180167A

Patent No. 6558950  
GENERAL INFORMATION:  
APPLICANT: Gordon C. Shore et al.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
FILE REFERENCE: 50013/004003  
CURRENT APPLICATION NUMBER: US/09/180,167A

```

: CURRENT FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: CA 2,158,988
: PRIOR FILING DATE: 1997-03-03
: PRIOR APPLICATION NUMBER: PCT/IB98/00706
: PRIOR FILING DATE: 1998-03-02
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: SeqIDSeq for Windows Version 4.0
: SEQ ID NO: 12
: LENGTH: 66
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-160-167A-12

```

Query Match	21.4%;	Score 80;	DB 4;	Length 66;
Best Local Similarity	34.5%;	Pred. No. 0.0086;		
Matches	20;	Conservative	11;	Mismatches 27;
			Indels	0;
			Gaps	0;

[illegible]

RESULT 6  
US-09-033-524B-12  
; Sequence 12, Application US/09033524B  
Default wa 500000

```

1  APPLICANT: Gordon C. Shore et al.
2  TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
3  TITLE OF INVENTION: APOPTOSIS
4  TITLE OF INVENTION: APOPTOSIS
5  FILE REFERENCE: 50013/004002
6  CURRENT APPLICATION NUMBER: US/09/033,524B
7  CURRENT FILING DATE: 1998-03-02
8  PRIOR APPLICATION NUMBER: CA 2,198,986
9  PRIOR FILING DATE: 1997-03-03
10 NUMBER OF SEQ ID NOS: 33
11 SOFTWARE: FastSeq for Windows Version 4.0
12 SEQ ID NO 12
13 LENGTH: 66
14 TYPE: PRT
15 ORGANISM: Homo sapiens
16 US-09-033-524B-12

```

Query Match	21.4%	Score 80;	DB 4;	Length 66;
Best Local Similarity	34.5%;	Pred. No. 0.0086;		
Matches	20;	Conservative	11;	Mismatches 27;
				Indels 0;
				Gaps 0;

Dy EIGELDQKSVSSLLFLMKYMGGRGKISKESFLDLVLELEKNLVAPDQLLEKCL 600  
::|::|::|::|::|::|::|::|  
Db DIGEQLDSDELAKLFLSDYIPQRKEPIKALMLFGLQEKMTLESNTSLFEKL 660

RESULT 7  
US-09-180-167A-31  
; Sequence 31, Application US/09180167A

```

1  APPLICANT: Gordon C. Shore et al.
2  TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
3  TITLE OF INVENTION: APOPOSIS
4  FILE REFERENCE: 50013/004003
5  CURRENT APPLICATION NUMBER: US/09/180,167A
6  PRIOR FILING DATE: 1999-09-10
7  PRIOR APPLICATION NUMBER: CA 2,198,988
8  PRIOR FILING DATE: 1997-03-03
9  PRIOR APPLICATION NUMBER: PCT/IB98/00706
10 PRIOR FILING DATE: 1998-03-02
11 NUMBER OF SEQ ID NOS: 33
12 SOFTWARE: FastSeq for Windows Version 4.0
13 SEQ ID NO 31
14 LENGTH: 64
15 TYPE: PRT
16 ORGANISM: Homo sapiens

```

US-09-180-167A-31

Query Match	18.5%;	Score 69;	DB 4;	Length 64;
Best Local Similarity	30.9%;	Pred. No. 0.19;		
Matches 17;	Conservative 11;	Mismatches 23;	Indels 4;	Gaps 1;

```

QY      3 EIGEDLDKSDVSSLIFLMKDYMGGRGKISKEKSFIDLVELEKINLVAPQDLLE 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     11 ELSEGISSENILKDMIFLLKDSLPK---TENTSLSPFLAFLEKQGKIDEDNLTGLE 61

```

RESULT 8  
US-09-033-524B-31

Sequence 31, Application US/09033524B  
Patent No. 6607880  
GENERAL INFORMATION:  
Additional Document: 31

APPLICANT: Gordon C. Shore et al.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
TITLE OF INVENTION: APOPTOSIS  
FILE REFERENCE: 50013/004002  
CURRENT APPLICATION NUMBER: US/09/033,524B  
ATTORNEY FILING DATE: 1999-03-02

PRIOR APPLICATION NUMBER: CA 2,198,988  
PRIOR FILING DATE: 1997-03-03  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 4.0

ORGANISM: Homo sapiens  
US-09-033-524B-31

Query Match	18.5%;	Score 65;	DB 4;	Length 64;
Best Local Similarity	30.9%;	Pred. No. 0.19;		
Matches 17;	Conservative 11;	Mismatches 23;	Indels 4;	Gaps 1;

```

QY      3 EIGEDLDKSDVSSLIFLMKDYMGGRGKISKEKSFDDLVELEKTNLVAPQDLLE 57
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      11 ELSEGISENLKDMIFLLKSLPK----TEMTSLFLAFLEQGKIDEDNLTCL 61

```

RESULT 9  
US-09-180-167A-10

Sequence 10, Application US/09180167A  
Patent No. 6558950  
GENERAL INFORMATION:  
APPLICANT: Gordon C. Shore et al.  
TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
TITLE OF INVENTION: APOPTOSIS

```

CURRENT APPLICATION NUMBER: US/09/180,167A
CURRENT FILING DATE: 1999-09-10
PRIORITY APPLICATION NUMBER: CA 2,198,988
PRIORITY FILING DATE: 1997-03-03
PRIORITY APPLICATION NUMBER: PCT/IB98/00706
PRIORITY FILING DATE: 1998-03-02
NUMBER OF SEQ. ID NOS: 33
SOFTWARE: Fasteq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 67
TYPE: PRT
ORGANISM: Homo sapiens
US-09-180-167A-10

```

Query Match	17.7%;	Score 66;	DB 4;	Length 67;
Best Local Similarity	29.8%;	Pred. No. 0.47;		
Matches	17;	Conservative	13;	Mismatches 27; Indels 0; Gaps 0;

```

Oy      4  IGEDIDKSDVSSLIFFLMKDYMGGKISIEKSFLLDVLVELEKLNLPAPDQDLLEKCL 600
      : | : : : | : : : : | : : : : | : : : : | : : : : |
Db      11 VSSSLSSSELTFLCLGRVGRKLERQSGDGLFMSMLBQNDEBERGHTTELLRELL 678

```

```
RESULT 10
US-09-033-524B-10
; Sequence 10, Application US/09033524B
; Patent No. 6607880
; GENERAL INFORMATION:
; APPLICANT: Gordon C. Shore et al.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
; FILE REFERENCE: 50013/004002
; CURRENT APPLICATION NUMBER: US/09/033,524B
; CURRENT FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: CA 2,198,988
; PRIOR FILING DATE: 1997-03-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-033-524B-10

Query Match          17.7%; Score 66; DB 4; Length 67;
Best Local Similarity 29.8%; Pred. No. 0.47;
Matches 17; Conservative 13; Mismatches 27; Indels 0; Gaps 0;

QY      4  IGEDLDKSDVSSLIIFLMKDYMGKISKESKESFLDLVLELEKLNVAPODLLEKCL 60
Db      11 VSSISSESELTETKFLCLGVRGKRLERVQSGDLDFVLMLEQNDLEPGHTLEIRELL 67

RESULT 11
US-09-823-266-15
; Sequence 15, Application US/09823266
; Patent No. 6613531
; GENERAL INFORMATION:
; APPLICANT: Burgess, Richard
; APPLICANT: Arthur, Terrance
; APPLICANT: Anthony, Larry
; APPLICANT: Bergendahl, Viet
; APPLICANT: Pietz, Bradley
; TITLE OF INVENTION: Sigma binding region of RNA polymerase and uses thereof
; FILE REFERENCE: 800.025U1
; CURRENT APPLICATION NUMBER: US/09/823,266
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,116
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 47
; TYPE: PRT
; ORGANISM: T. maritima
US-09-823-266-15

Query Match          13.8%; Score 51.5; DB 4; Length 47;
Best Local Similarity 38.9%; Pred. No. 18;
Matches 14; Conservative 6; Mismatches 11; Indels 5; Gaps 1;

QY      6  EDLDKSDVSSLIIFLMKDYMGKISKESKESFLDLVLE 41
Db      8  EKLTISNLRIVASIAKRYGGRG-----LSFDLIQE 38

RESULT 12
US-09-823-266-13
; Sequence 13, Application US/09823266
; Patent No. 6613531
; GENERAL INFORMATION:
; APPLICANT: Burgess, Richard
; APPLICANT: Arthur, Terrance
; APPLICANT: Anthony, Larry
; APPLICANT: Bergendahl, Viet
; APPLICANT: Pietz, Bradley

RESULT 13
US-09-180-167A-11
; Sequence 11, Application US/09180167A
; Patent No. 6558950
; GENERAL INFORMATION:
; APPLICANT: Gordon C. Shore et al.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
; FILE REFERENCE: 50013/004003
; CURRENT APPLICATION NUMBER: US/09/180,167A
; CURRENT FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: CA 2,198,988
; PRIOR FILING DATE: 1997-03-03
; PRIOR APPLICATION NUMBER: PCT/IB98/00706
; PRIOR FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-180-167A-11

Query Match          13.7%; Score 51; DB 4; Length 67;
Best Local Similarity 28.3%; Pred. No. 33;
Matches 15; Conservative 12; Mismatches 26; Indels 0; Gaps 0;

QY      8  LDKSDVSSLIIFLMKDYMGKISKESKESFLDLVLELEKLNVAPODLLEKCL 60
Db      15  LSGNDLMEIKFLCRERVSKRLERVQSGDLDFVLMLEQNDLEPGHTLEIRELL 67

RESULT 14
US-09-033-524B-11
; Sequence 11, Application US/09033524B
; Patent No. 6607880
; GENERAL INFORMATION:
; APPLICANT: Gordon C. Shore et al.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING
; FILE REFERENCE: 50013/004002
; CURRENT APPLICATION NUMBER: US/09/033,524B
; CURRENT FILING DATE: 1998-03-02
; PRIOR APPLICATION NUMBER: CA 2,198,988
; PRIOR FILING DATE: 1997-03-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 67
; TYPE: PRT
```



ORGANISM: Mus musculus  
US-09-033-524B-11

US-09-033-524B-11

Query Match	13.7%	Score 51;	DB 4;	Length 67;
Best Local Similarity	28.3%	Pred. No. 33;		
Matches 15; Conservative	12;	Mismatches	26;	Indels 0;
				Gaps 0;

Matches 15; Conservative 12; Mismatches

Matches	15;	Conservative	12;	Mismatches	26;	Indels	0;	Gaps	0;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

DQ 8 LPSDVSSLIPLMKDYMGRGRTSKEKSFLLDLVELEKTLNVAPOGLLLEKCL 60  
 : : : : : : : : : : : : : :  
 Db 15 LSGNDLMELKFCREVRVSKRLERVQSGIDLFTVLIEQNLDLERGHGTGLRELTL 67

Db 15 LSGNDLMEIKFLCRERVSKRKLERVQSGDLFTVLLEQNDLERGHTGLRELL 67

RESULT 15

US-09-621-976-7353

; Sequence 7353, Application US/09621976

Patent No. 6639063

GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Robert S

APPLICANT: Jober, S.  
APPLICANT: Giordano

APPLICANT: GIORDANO, J. J.  
TITLE OF INVENTION: ESTS

FILE REFERENCE: GENSET.054PR2

```

/ ***** SENDIN : 001 *****
; CURRENT APPLICATION NUMBER: U

```

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 7353

; LENGTH: 69

TYPE: PRT

ORGANISM: Homo sapiens

US-09-621-976-7353

Query Match

Query Match	13.5%;	Score 50.5;	DB 4;	Length 69;
Best Local Similarity	31.4%;	Pred NC	40.	

Best Local Similarity 31.4%; Pred. No. 40; Matches 22: Conservative 13: Mismatch

Matches	22;	conservative	13;	Mismatches	22;	Indels	13;	Gaps	5;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 8 LDKSDVSSLLFLMKDYNRGKSKSEKSFLLDLVELEKLNLVAPDDLLLEKCNIRID 67  
:: :: :: ||| :: | : : :  
Db 1 MDNSNIWAM-TEMDSGSGRTI----SFVQYEALKTGLCTEDE-DLQD-----DGKHIT 50

Db 1 MDNSNIAM-FEMDSSGRGTI-----SFVQKEALKTLGLCTEDE-DLQD-----DGHKIT 50

QY 68 L---KTKIQK 74

Db 51 LDKFKEVNK 60

Search completed: September 9, 2004, 05:54:39  
Job time : 34 secs

Job time : 34 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:48:31; Search time 40 Seconds

(without alignments)  
180.359 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75

Sequence: 1 MAEIGEDLDKSDVSLIFLM.....LEKLNKHIDUKTKIQK 75

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 23401

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	15.0	60	2	139957
2	53.5	14.3	59	2	C36493
3	53	14.2	48	2	H83759
4	50.5	13.5	66	2	AE3493
5	50.5	13.5	68	2	FE3799
6	49.5	13.3	70	2	BE4497
7	48.5	13.0	66	2	G1834
8	48.5	13.0	66	2	T00821
9	47.5	12.7	62	2	T12832
10	47	12.6	40	2	S45688
11	46	12.3	46	2	C41372
12	45.5	12.2	44	2	F70225
13	45.5	12.2	58	2	DE9871
14	45.5	12.2	60	2	T12910
15	45	12.1	45	2	B43859
16	45	12.1	65	2	A34945
17	45	12.1	62	2	S77797
18	45	12.1	62	2	C90761
19	45	12.1	62	2	H55624
20	45	12.1	72	2	B67705
21	45	12.1	74	2	BE4127
22	44.5	11.9	66	2	G64683
23	44.5	11.9	66	2	G64683
24	44.5	11.9	71	2	T07852
25	44	11.8	66	2	T12638
26	44	11.8	71	2	A62362
27	43.5	11.7	30	2	S34761
28	43.5	11.7	64	2	T39013
29	43.5	11.7	64	2	T39013

30	43	11.5	70	2	S77893	hypothetical prote
31	42.5	11.4	65	2	C82339	hypothetical prote
32	42.5	11.4	66	2	C97870	oxidoreductase hom
33	42.5	11.4	74	2	E90827	hypothetical prote
34	42.5	11.4	74	2	C96547	hypothetical prote
35	42	11.3	57	2	C97758	hypothetical prote
36	42	11.3	65	2	T03652	probable carbonate
37	42	11.3	70	2	S48855	DNA-directed RNA p
38	41.5	11.1	65	2	C82290	hypothetical prote
39	41.5	11.1	67	2	H96936	probable metal-bin
40	41.5	11.1	72	2	G72399	ribosomal protein
41	41.5	11.1	73	2	E95078	conserved hypochet
42	41.5	11.1	75	2	AB1245	exodeoxyribonuclea
43	41.5	11.1	75	2	AF1607	exodeoxyribonuclea
44	41	11.0	40	2	T07583	hypothetical prote
45	41	11.0	50	2	H81250	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

139957  
degradative enzyme production factor degr - Bacillus subtilis

C/Species: Bacillus subtilis  
C/Date: 19-Jul-1996 #sequence revision 19-Jul-1996 #text\_change 15-Oct-1999

C/Accession: 139957; I39958; A69614

R/Nagami, Y.; Tanaka, T.

U. Bacteriol. 166, 20-28, 1986

A/Title: Molecular cloning and nucleotide sequence of a DNA fragment from Bacillus nat

A/Reference number: 139957; MUID:86168015; PMID:3082853

A/Accession: 139957

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-60 <RES>

A/Cross-References: GB:M12917; NID:g143354; PIDN:AAA22671.1; PID:g143355

R.Yang, M.; Shimoto, H.; Ferrati, E.; Henner, D.U.

J. Bacteriol. 169, 434-437, 1987

A/Title: Characterization and mapping of the Bacillus subtilis ptrr gene.

A/Reference number: 139958; MUID:87083406; PMID:3098734

A/Accession: 139958

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-60 <RES>

A/Cross-References: GB:M15318; NID:g143356; PIDN:AAA22672.1; PID:g143357

R.Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Berte

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.T.; Errington, U.; Fader, C.; Ferrati, E

Nature 350, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F

Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y. M.; Ogawa, K.; Ogihara, B.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon

A/Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, U.; Sekowska, A.; Sero

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama

T.; Whalers, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,

A/Authors: Yoshikawa, H.F.; Zunshtein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: A69614

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-60 <KUN>

A/Cross-References: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14112.1; PID:el183641

A/Experimental source: strain 168

A/Genes: degr

Query Match 15.0%; Score 56; DB 2; Length 60;  
Best Local Similarity 27.8%; Pred. No. 40;  
Matches 15; Conservative 14; Mismatches 17; Indels 8; Gaps 1;

Oy      28 KISKESKPLDVELEKLNTVAPQDLDEK-----CLKNHRIIDLKRIQ 73  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db      7 KLILHKFIETIYSDLEEELADIAKKKGKPSMEKVXEIEBRCCKNLTALIEIQMK 60

RESULT 2  
C36493  
H+-transporting two-sector ATPase (EC 3.6.3.14) epsilon chain - Sulfolobus acidocaldarius  
C/Species: Sulfolobus acidocaldarius  
C/Date: 12-Apr-1991 #sequence\_revision 12-Apr-1991 #text\_change 03-Jun-2002  
C/Acession: C36493  
R/Denda: K.; Konishi, J.; Hajiro, K.; Oshima, T.; Date, T.; Yoshida, M.  
T./BIOL. Chem. 265, 24507-24513, 1990  
A./Title: Structure of an ATPase operon of an acidothermophilic archaeobacterium, Sulfolob.

A./Reference number: A36493; MUID:91072342; PMID:2147683  
A./Accession: C36493  
A./Status: preliminary  
A./Molecule type: DNA  
A./Residues: 1-59 <DEN>  
A./Cross-references: GB:M57238; GB:U05671; NID:g152915; PID:g152918  
C/keywords: hydrolase

Query Match	14.3%;	Score 53.5;	DB 2;	Length 59;
Best Local Similarity	30.6%;	Pred. No. 70;		
Matches 22;	Conservative 14;	Mismatches 17;	Indels 19;	Gaps 5;

Dz  
Qy

7 DLDKSDVSSLIFLMKDVMGRGKISKEKSFLLDVLVELEKNTL-----VAPDQDLBLEKCLXN 62  
::||| : ||: | | | |  
3 EIDKSTIDKYINILK-----SKLDQCKNEL-----LSKINMEYEKTLQRDLBELEKLGN 52

QY	63	IHRIDLKTKIQK	74
Db	53	I-----LK-EVQK	59

RESULT 3  
H83759  
hypothetical protein BH0680 [imported] - Bacillus halodurans (strain C-125)  
C|Species: Bacillus halodurans  
C|Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C|Accession: H83759  
R|Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirose, T.  
Nucleic Acids Res. 28, 4317-4331, 2000  
A|Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A|Reference number: A83650; MUID:20512582; PMID:11058132  
A|Accession: H83759  
A|Status: preliminary  
A|Molecule type: DNA  
A|Residues: 1-48 <STO>  
A|Cross-references: GB:BA001510; GB:BA000004; NID:g10173440; PIDD:BA04599.1; GSPDB:GN000000000  
A|Experimental source: strain C-125  
C|Genetics:  
/Gene: BH0680

Query MarC... 14.2% Score 53; DB 2; Length 48;  
Best Local Similarity 41.3%; Pred. No. 63;  
Matches 19; Conservative 6; Mismatches 11; Indels 10; Gaps 2

Oy 24 MNRGKISKEKSFVLYVELEKLNVLAPQDULREKLCKNHRDLK 69  
| | | | | | | | | | | | | | | | | | | |  
Db 1 MGR---KLSFEELVENKEFLNDPQLSKIK-----RDEK 36  
| | | | | | | | | | | | | | | | | | | |

```

RESULT 4
AF3493
Hypochemical cytosolic protein EME11932 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3493
R:DelVecchio, V.G., Kaparatel, V., Redker, R.U., Patra, G., Mujer, C., Los, T., Iancova,
I., Mazur, M., Goldsman, E., Selkov, E., Elzert, P.H., Hegius, S., O'Callaghan, D., Leclerc,
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

```

A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3493  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-66 <KUS>  
A:Cross-references: GB:AE009917; PID:g17983979; GSPDB:GN00150  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME11932  
A:Map position: 1

```

Oy      24 MEGKISKESKSFDDLEVEL-----EKNLV----APQDLDLEKCL 60
        | | : : : || | | | | | | | | | |
Db      1 MORYFDDKRILTEAIVELLNGIEPEDDLDVLSRIGPVDLDLMQEC 48

```

RESULT 5  
F69799  
Hypothetical protein yezA - Bacillus subtilis  
Bacillus subtilis subsp. spizizenii

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C;Accession: F69799

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; C.; Enllich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997

Aranovich, S., Hershkovitz, E., Hershkovitz, E., Hosono, S., Hall, M., Iech, J., Harwood, C.R., Henaut, A., Hilbert, H., Hollappel, S., Koellers, P., Königstein, G., Krogh, S., Kunano, M., Kurita, K., Kutter, T., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Authors: lauber, j.; lazarevic, v.; lee, s.m.; levine, a.; liu, h.; masuda, s.; mau

Y. M., Ogata, K., Gwigira, A., Uudea, B., Pak, S.H., Patis, V., Poni, I., T. Forecella, R., Rieger, M., Rivolla, C., Rocha, E., Roche, B., Rose, M., Sadle, Y., Sato, T., Schall, A., Authors: Schiechl, S., Schroeder, R., Scetione, F., Sekiguchi, J., Sekuska, A., Seya, A., Kuchni, M., Tamakoshi, A., Tanaka, T., Terestera, P., Terebina, A., Tononi, A., Tosato, V., Uchiyama, A.

T. Winters, P. Wipat, A. Yamamoto, E. Yoshikawa, H. F. Danchin, A. A. Authors: Yoshikawa, H. F. Danchin, A. A. Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* strain 9804/033. PMID: 984372

A:Accession: F69799  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Accession numbers: GB:Z99107, GB:AL009126, MID:S2632866, PIDN:CA812500.1, PID:el1826  
 A:Cross-references: GB:Z99107, GB:AL009126, MID:S2632866, PIDN:CA812500.1, PID:el1826  
 A:Experimental source: strain 168  
 C:Genetics:

Query Match	13.5%	Score	50.5;	DB 2;	Length	68;
Best local similarity	39.2%	Pred. No.	1.7e+02;			
A/Gene: yezA						

```
Qy      8 LDKSDVSSLIPLMKDYMGRGKISKKE--SFIDLVELEKLNVAPOQLDL 56
          :| | | | :| | | | | | | | | | | | | | | |
Db      1 MDKLREELLFLVKELM-NPTLDEKVSLEYDL---LEK-NVPYPAPSDLI 46
```

RESULT 6  
B64497  
hypochemical protein MJ1579 - Methanococcus jannaschii  
C1species: Methanococcus jannaschii  
C1date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999  
C1accession: B64497  
R1Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak  
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinschenk, K.G.; Merrick, J.M.; Glodex,  
ron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurest, M.A.  
Science 273, 1058-1073, 1996  
A1authors: Kaine, B.P.; Borodovsky, M.; Kleck, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
A1title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

RESULT 10  
S45688





*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:48:31 / Search time 23 Seconds  
(without alignments)  
169.794 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75  
Perfect score: 373  
Sequence: 1 MAFIGEDLDKSDVSLIFLM.....LEKCLKNHRIIDUKTIQKY 75

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 8856

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	15.0	60	1	DEGR_BACCU
2	53.5	14.3	59	1	MPPE_SUTRO
3	49.5	13.3	70	1	YF79_METJA
4	48.5	13.0	66	1	RL29_HELPT
5	48	12.9	52	1	YOMI_HAEIN
6	45	12.1	60	1	SENN_BACNA
7	45	12.1	60	1	YG9A_CLOTE
8	44.5	11.9	64	1	RL29_LACPL
9	44.5	11.9	66	1	RL29_HELPT
10	44.5	11.9	71	1	RPRX_BRANA
11	44	11.8	71	1	R316_TANAP
12	44	11.8	73	1	YAEI_SCHPO
13	43.5	11.7	30	1	SDHA_CLOPR
14	43.5	11.7	64	1	YDF6_SCHPO
15	43	11.5	65	1	SENS_BACCU
16	42	11.3	70	1	RPRX_YEAST
17	42	11.3	71	1	RL30_BORBU
18	41.5	11.1	70	1	RL28_THERA
19	41.5	11.1	75	1	EXYS_LISTN
20	41.5	11.1	75	1	EXYS_LISTN
21	41	11.0	75	1	EXYS_LISTN
22	40.5	10.9	58	1	YQZD_THEN
23	39.5	10.6	63	1	RL29_YERPE
24	39.5	10.6	65	1	RL29_YERPE
25	39.5	10.6	66	1	Y13G_BPTA
26	39.5	10.6	65	1	RL29_YERPE
27	39	10.5	75	1	RS21_UREPA
28	39	10.5	65	1	RL35_BUCAP
29	39	10.5	67	1	YB00_METJA
30	39	10.5	73	1	Y179_URBPA
31	38.5	10.3	63	1	RPO9_UCVU
32	38.5	10.3	64	1	DN72_SULAC
33	38.5	10.3	66	1	RK33_OENHO

34	38.5	10.3	74	1	VP2_SSV1	P20224 sulfolobus
35	38	10.2	42	1	PSA1_PICGL	O64096 picea glauc
36	38	10.2	49	1	VRD1_BACCU	O07946 bacillus vi
37	38	10.2	65	1	VC13_YACCC	P21044 vaccinia vi
38	38	10.2	66	1	COP1_HELPT	Q48271 helicobacte
39	38	10.2	68	1	YF02_ARCTU	O28770 archaeoglob
40	38	10.2	70	1	BTB7_MYCTU	O05845 mycobacteri
41	38	10.2	70	1	PSA6_CVAPA	P48114 cyanophora
42	37.5	10.1	53	1	YCX4_GUTH	O78452 guillardia
43	37.5	10.1	56	1	VGK_BPPX	P03653 bacterioph
44	37.5	10.1	65	1	DN71_SULAC	P13123 sulfolobus
45	37.5	10.1	73	1	ACP_LACUA	Q9cmf9 lactococcus

## ALIGNMENTS

RESULT 1	ID	DEGR_BACCU	STANDARD	PRT	60 AA
AC	P06563				
DT	01-JAN-1988 (Rel. 06, Created)				
DT	01-JAN-1988 (Rel. 06, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Regulatory protein degr.				
GN	DEGR OR PRTR OR BSU21940				
OS	Bacillus subtilis, and				
OS	Bacillus subtilis var. natto.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423, 86029;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=B. subtilis;				
RC	MEDLINE=96349105; PubMed=8760912;				
RA	Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,				
RA	Seror P.,				
RT	"Sequence analysis of the Bacillus subtilis chromosome region between				
RT	the <i>serA</i> and <i>kds</i> loci cloned in a yeast artificial chromosome.";				
RL	Microbiology 142:2005-2016(1996).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=B. subtilis; STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bertero M.G., Besieres P., Bortolin A., Borchert S.,				
RA	Borries R., Boursier L., Brans A., Braun W., Bignelli S.C., Bron S.,				
RA	Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,				
RA	Choi S.K., Codani U.J., Conerton I.F., Cummings N.J., Daniel R.A.,				
RA	Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,				
RA	Enlian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,				
RA	Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grand G.,				
RA	Guiseppi G., Guy B.J., Haga K., Haele K., Haele K., Haele K.,				
RA	Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,				
RA	Joris B., Karamata D., Kasehara Y., Kaefer-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koster P., Koningsstein G., Kroh S., Kumano M.,				
RA	Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,				
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,				
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,				
RA	Parro V., Pohl T.M., Portetelle D., Portwilk S., Prescott A.M.,				
RA	Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,				
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,				
RA	Sato T., Scanlan E., Schleich S., Schroeter F., Scoffone F.,				
RA	Sekiguchi U., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,				
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,				

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandendol M., Vannier P., Vassarotti A.,  
RA Viari A., Wandt R., Wedler E., Wedler H., Wetzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT subtilis";  
RL Nature 390:249-256(1997).  
RN  
RP  
R [4]  
R SEQUENCE FROM N.A.  
R SPECIES=B.s.natto;  
R MEDLINE=86168015; PubMed=3082853;  
RA Nagami Y., Tanaka T.;  
RT "Molecular cloning and nucleotide sequence of a DNA fragment from  
RT *Bacillus natto* that enhances production of extracellular proteases  
RT and levanucrase in *Bacillus subtilis*.";  
RL J. Bacteriol. 166:20-28(1986).  
CC -I- FUNCTION: LEADS, IN B.SUBTILIS, TO ENHANCED PRODUCTION OF  
CC LEVANSUCRASE, ALKALINE PROTEASE, AND NEUTRAL PROTEASE. IN B.NATTO  
CC IT IS NONESSENTIAL FOR GROWTH OR EXPRESSION OF PROTEASES AND  
CC LEVANSUCRASE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M15318; AAA22672.1; -;  
DR EMBL; L77246; AAA96622.1; -;  
DR EMBL; Z99115; CAB14112.1; -;  
DR EMBL; M12917; AAA22671.1; -;  
DR PIR; I39957; I39957.  
DR Subtilisin; Bg10659; degr.  
KW Transcription regulation; Complete proteome.  
SQ SEQUENCE 60 AA; 7109 MW; C7B7F892FF0AB131 CRC64;  
  
Query Match 15.0%; Score 56; DB 1; Length 60;  
Best Local Similarity 27.8%; Pred. No. 29;  
Matches 15; Conservative 14; Mismatches 17; Indels 8; Gaps 1;  
  
QY 28 KISKESFDDLVLEKLNIVAPDLDLEK-----CLKHRIIDLKTIQK 73  
Db 7 KLIIHKTFFIIVSDLEELDIADKGRPSMEKVEIEIQRKONILAIIEIQMKIK 60  
  
RESULT 2  
ID TYPE SULTO STANDARD; PRT; 59 AA.  
AC P23039;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Membrane-associated Arpase epsilon chain (EC 3.6.3.14) (SUL-ATPase  
DE epsilon)  
GN ARPE OR ST1438.1 OR STS172.  
OS Sulfolobus tokodaii, and  
OS Sulfolobus acidocaldarius.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=11955, 2285;  
RN  
RP  
R SEQUENCE FROM N.A.  
R SPECIES=S.acidocaldarius;  
R MEDLINE=91072342; PubMed=2147683;  
RA Denda K., Konishi J., Hajiro K., Oshima T., Date T., Yoshida M.,  
RT "Structure of an Arpase operon of an acidothermophilic  
RT archaebacterium, *Sulfolobus acidocaldarius*,"  
RL J. Biol. Chem. 265:21509-21513(1990).  
RN  
RP  
R SEQUENCE FROM N.A.

RC SPECIES=S.tokodaii; STRAIN=JCM 10545 / 7;  
RX MEDLINE=2156156; PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kusida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagis M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.,  
RT "Complete genome sequence of an aerobic thermacidophilic  
RT Crenarchaeon, *Sulfolobus tokodaii* strain 7.";  
RL DNA Res. 8:123-140(2001).  
CC -I- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +  
CC H(+) (out).  
CC  
CC -I- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE?) SUBUNITS.  
CC ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.  
CC -I- SIMILARITY: SOME, TO E.HIRAE NTPA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M57238; AAA72942.1; -;  
DR EMBL; AP000986; BAB6506.1; -;  
DR PIR; C36493; C36493.  
KW Hydrolyase; Hydrogen ion transport; Complete proteome.  
SQ SEQUENCE 59 AA; 7038 MW; 4C07944C73225D4 CRC64;  
  
Query Match 14.3%; Score 53.5; DB 1; Length 59;  
Best Local Similarity 30.6%; Pred. No. 51;  
Matches 22; Conservative 14; Mismatches 17; Indels 19; Gaps 5;  
  
QY 7 DLDKSDVSLIFIMDYGRKISKESFDDLVLEKLN-----VAPDLDLEKLN 62  
Db 3 ELDKSTIKRYINILK-----SKDPQKNEL-----LSKINNEYKTLKQRLDELEKLN 52  
  
QY 63 IHRIDLKTIQK 74  
Db 53 I-----LK-EVQK 59  
  
RESULT 3  
ID TYPE METUA STANDARD; PRT; 70 AA.  
AC Q58974;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M1579.  
GN M1579.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN  
RP  
R SEQUENCE FROM N.A.  
R STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
RA Utecherak T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-F., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,  
RT "Complete genome sequence of the mechanogenic archaeon, *Methanococcus*  
RT jannaschii,"  
RL Science 273:1058-1073(1996).  
RN  
RP  
R SEQUENCE FROM N.A.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67598; AAB99607.1; -.
DR PIR; B64497; B64497.
DR TIGR; M01579; -.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 53 LYS-RICH.
SQ SEQUENCE 70 AA; 8483 MW; A485930C4F232461 CRC64;

Query Match 13.3%; Score 49.5; DB 1; Length 70;
Best Local Similarity 29.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 12; Mismatches 29; Indels 5; Gaps 2;

QY 10 KSDVSLFLMKDYGRGKISKESFIDLVELEKLNIVAPDQDLLEKCKIHRIDLK 69
DB 4 REIELLNFDFVLPYANNEFIKE--LLESVNSLELE---QKVELLEKETDITKTDLK 58
DB 70 TKIQ 74
DB 59 ILLEK 63

RESULT 4
RL29 HELPU STANDARD; PRT; 66 AA.
AC 09ZJ51;
ID RL29 HELPU STANDARD; PRT; 66 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L29.
GN RPMC OR JHP1231.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
CC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
CC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Weiberg D., Mills S.D., Chang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- SIMILARITY: Belongs to the L29P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001547; AAD06797.1; -.
DR PIR; G71834; G71834.
DR HAMAP; MF 00374; -.
DR InterPro; IPR001854; Ribosomal_L29.
DR Pfam; PF00831; Ribosomal_L29; 1.
DR TIGRfam; TIGR00012; L29; 1.
DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 66 AA; 7752 MW; 959E77507F15940E CRC64;

Query Match 13.0%; Score 48.5; DB 1; Length 66;

```

```

Best Local Similarity 30.4%; Pred. No. 1.8e+02;
Matches 14; Conservative 12; Mismatches 15; Indels 5; Gaps 2;

QY 30 SKESFLDLYVELEKLNIVAPDQDLLEKCKIHRIDLTKIQY 75
DB 20 AKKAEFLRKYKMQMQLSNPNE--IKARRNIRIN--TAINAY 60

RESULT 5
YQAH HAEIN STANDARD; PRT; 52 AA.
AC P56507; O86238;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H11434.2.
GN H11434.2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fire L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION.
RX MEDLINE=98248213; PubMed=9588799;
RA Rudd K.E., Humphrey-Smith I., Wasinger V.C., Bairoch A.;
RT "Low molecular weight proteins: a challenge for post-genomic
RT research.";
RL Electrophoresis 19:536-544(1998).
CC -1- SIMILARITY: Belongs to the UPF0181 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32822; AAC23083.1; ALT_INIT.
DR TIGR; H11434.2; -.
DR HAMAP; MF 00507; -.
DR InterPro; IPR005371; UPF0181.
DR Pfam; PF03701; UPF0181; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 52 AA; 5903 MW; FFB836E185FB262 CRC64;

Query Match 12.9%; Score 48; DB 1; Length 52;
Best Local Similarity 22.5%; Pred. No. 1.6e+02;
Matches 9; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

QY 31 KEKSLDLYVELEKLNIVAPDQDLLEKCKIHRIDLTK 70
DB 10 QQKAVQEIQLKMQGISGEAIQIVAKALREIHKDKXT 49

RESULT 6
SENN_BACNA

```

```

ID  SENN_BACNA      STANDARD;      PRT;      60 AA.
AC  P21666;
DT  01-MAY-1991 (Rel. 18, Created)
DT  01-MAY-1991 (Rel. 18, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  Transcriptional regulatory protein senn.
GN  SENN.
OS  Bacillus subtilis var. natto.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=86029;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=90063530; PubMed=3269394;
RA  Wong S.-L., Wang L.-F., Doi R.H.;
RT  "Cloning and nucleotide sequence of senn, a novel 'Bacillus natto'
RT  (B. subtilis) gene that regulates expression of extracellular protein
RT  genes.";
RL  J. Gen. Microbiol. 134:3269-3276(1988).
CC  -1- FUNCTION: REGULATES THE EXPRESSION OF EXTRACELLULAR-PROTEIN GENES
CC  -1- OF BACILLUS NATTO.
CC  -1- SIMILARITY: TO SEVERAL B.SUBTILIS RNA POLYMERASE SIGMA FACTORS.
CC  -1- SIMILARITY: TO B.SUBTILIS SENS.
DR  PIR: A34945; A34945.
RW  Transcription regulation; DNA-binding.
FT  DNA BIND 11 31 H-R-H MOTIF (BY SIMILARITY).
SQ  SEQUENCE 60 AA; 7220 MM; 0CB30106C062B6F6 CRC64;

Query Match 12.1%; Score 45; DB 1; Length 60;
Best Local Similarity 26.0%; Pred. No. 3.7e+02;
Matches 13; Conservative 12; Mismatches 17; Indels 8; Gaps 2;

QY 20 MKQYMRGKSKSEKSF-----LDLVELEKLNVAPODLLEKLNRI 63
DB 3 VKRKGKRRKRRKRTGNGQILPLELLEIKKKRII--NSAEIMEIYVKI 50

RESULT 7
Y99F_CLOTE      STANDARD;      PRT;      60 AA.
AC  P60083;
DT  15-MAR-2004 (Rel. 43, Created)
DT  15-MAR-2004 (Rel. 43, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  Hypothetical UPF0291 protein CTC01690.1.
GN  CTC01690.1.
OS  Clostridium tectani.
OC  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC  Clostridium.
OX  NCBI_TaxID=1513;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=Massachusetts / B88;
RA  MEDLINE=22457253; PubMed=12552129;
RA  Brueggemann H., Baumeister S., Fritzsche W.F., Wiesner A., Liesegang H.,
RA  Decker I., Herzberg C., Martiner-Arias R., Merkl R., Henne A.,
RA  Gottschalk G.;
RT  "The genome sequence of Clostridium tectani, the causative agent of
RT  tetanus disease.";
RL  Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Proteolysis).
CC  -1- SIMILARITY: Belongs to the UPF0291 family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AE015941; -; NOT ANNOTATED CDS.
RW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 60 AA; 7133 MM; 3A607AE48063505 CRC64;

```

```

Query Match 12.1%; Score 45; DB 1; Length 60;
Best Local Similarity 34.8%; Pred. No. 3.7e+02;
Matches 16; Conservative 10; Mismatches 10; Indels 10; Gaps 3;

QY 28 KISKSEKFLDVELEKLNVAPODLLEKLNRIHRIIDKTKIQ 73
DB 15 KSKSEKGE-----LIEBEKI-----EQAEIREKLYKNI--RSNFRAQLE 50

RESULT 8
RL29_LACPL      STANDARD;      PRT;      64 AA.
AC  O88XX8;
DT  15-MAR-2004 (Rel. 43, Created)
DT  15-MAR-2004 (Rel. 43, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  50S ribosomal protein L29.
GN  RPMC OR LP_1043.
OS  Lactobacillus plantarum.
OC  Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC  Lactobacillus.
OX  NCBI_TaxID=1590;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=NCIMB 8826 / WCFS1;
RX  MEDLINE=22480296; PubMed=12566566;
RA  Kleerebezem M., Boekhorst U., van Kranenburg R., Molenaar D.,
RA  Kluipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA  Fiers M.W.B.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA  Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA  De Vos W.M., Sierzen R.J.;
RT  "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL  Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
CC  -1- SIMILARITY: Belongs to the L29 family of ribosomal proteins.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AL935254; CAD63579.1; -
DR  HAMAP: MF_00374; -; 1
DR  InterPro: IPR001854; Ribosomal_L29.
DR  Pfam: PF00831; Ribosomal_L29; 1.
DR  PROSITE: PS00579; RIBOSOMAL_L29; 1.
RW  Ribosomal protein; Complete proteome.
SQ  SEQUENCE 64 AA; 7538 MM; 2357F52F5704033F CRC64;

Query Match 11.9%; Score 44.5; DB 1; Length 64;
Best Local Similarity 36.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 9; Mismatches 11; Indels 13; Gaps 4;

QY 29 ISKSEKFLDVELEKLNVAPODLLEKLNRIHRIIDKTKIQ 74
DB 15 LEKSEKSYD-----ELFVLRFQALNGOLENTARLRKQYKXNIARI--KTALRE 59

RESULT 9
RL29_HELIPY      STANDARD;      PRT;      66 AA.
AC  P56052;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  50S ribosomal protein L29.
GN  RPMC OR HPI311.
OS  Helicobacter pylori (Campylobacter pylori).
OC  Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC  Helicobacteraceae; Helicobacter.

```



```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003596; BAB76150.1; -.
DR PIR; AC2362; AC2362.
DR HAVAP; MF_00270; -.
DR InterPro; IPR001648; Ribosomal_S18.
DR Pfam; PF01084; Ribosomal_S18; 1.
DR PRINTS; PR00974; RIBOSOMLS18.
DR ProDom; PD002239; Ribosomal_S18; 1.
DR TIGRPMs; TIGR00165; S18; 1.
DR PROSITE; PS00057; RIBOSOMAL_S18; 1.
DR Ribosomal protein; RNA-binding; rRNA-binding; Complete proteome.
SQ SEQUENCE 71 AA; 8282 MW; A193360B57FFBC1 CRC64;

Query Match 11.8%; Score 44; DB 1; Length 71;
Best Local Similarity 29.1%; Pred. No. 5.5e+02;
Matches 16; Conservative 8; Mismatches 19; Indels 12; Gaps 2;

QY 5 GEDLDKSDVSLIFLMKDYNGRGI-----SKESFLDLYVELEKLNIVAP 50
Db 14 GEPIDYXVD---LTKRFTFRGKILPRITGTLAKQRELTALIKSRVLALP 65

RESULT 12
ID YDTE_SCHPO STANDARD; PRT; 73 AA.
AC Q10167;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C26A3.14c in chromosome I.
GN SPAC26A3.14c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Sgouros K., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
James K., Jones L., Jones M., Leather S., McDonald S., Mealen J.,
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitch E.,
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
Woodward J., Voiclaert G., Aert R., Robben J., Grymoprez B.,
Wetjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,
Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
Lucas M., Rochet M., Gallard C., Tallada V.A., Garcon A., Thode G.,
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
Domiguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
Cernutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
The genome sequence of Schizosaccharomyces pombe.
Nature 415:871-880(2002).
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69240; CAA93237.1; -.
DR PIR; T38402; T38402.
DR GeneDB; SPCmbc; SPAC26A3.14c; -.
DR Hypothetical protein.
SQ SEQUENCE 73 AA; 8249 MW; B56C52B7B3435411 CRC64;

Query Match 11.8%; Score 44; DB 1; Length 73;
Best Local Similarity 43.8%; Pred. No. 5.7e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 60 LKNIHRIDPKTKIQKY 75
Db 34 VKKINHDYKTAVEKY 49

RESULT 13
ID SDHA_CLOPR STANDARD; PRT; 30 AA.
AC P80212;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-serine dehydratase, alpha chain (EC 4.3.1.1) (L-serine deaminase)
DE (SDH) (L-SD) (Fragment).
OS Clostridium propionicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=28446;
[1]
RN SEQUENCE.
RP STRAIN=ATCC 25522;
RX MEDLINE=93345521; PubMed=8344301;
Hofmeister A.E.M., Grabowski R., Linder D., Buckel W.;
"l-serine and l-threonine dehydratase from Clostridium propionicum.
Two enzymes with different prosthetic groups."
RT Eur. J. Biochem. 215:341-349(1993).
RL -1- CATALYTIC ACTIVITY: L-serine = pyruvate + NH(3).
CC -1- COFACTOR: 4Fe-4S cluster.
CC -1- PATHWAY: Gluconeogenesis from serine.
CC -1- SUBUNIT: Heterodimer of an alpha chain and a beta chain.
CC -1- SIMILARITY: Belongs to the iron-sulfur dependent L-serine
CC dehydratase family.
DR PIR; S34761; S34761.
KW Lyase; iron-sulfur; 4Fe-4S; Gluconeogenesis.
FT NON TER 30
SQ SEQUENCE 30 AA; 3390 MW; 2AA4843780234641 CRC64;

Query Match 11.7%; Score 43.5; DB 1; Length 30;
Best Local Similarity 41.4%; Pred. No. 2.6e+02;
Matches 12; Conservative 3; Mismatches 9; Indels 5; Gaps 1;

QY 31 KKSFLDLYVELEKLN-----LVAPDQID 54
Db 2 KYDSLADLYVQAEKQNVPLXLIKXQDQE 30

RESULT 14
ID YDTE_SCHPO STANDARD; PRT; 64 AA.
AC Q14211;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C6B12.06c in chromosome I.
GN SPAC6B12.06c.
```

RA MEDLINE=90202712; PubMed=2108127; PubMed=9202460;  
 RA Mang L.-F., Doi R.H.;  
 RT "Complex character of sens, a novel gene regulating expression of  
 RT extracellular-protein genes of *Bacillus subtilis*."  
 RT J. Bacteriol. 172:1939-1947 (1990).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=168;  
 RC MEDLINE=97346037; PubMed=9202460;  
 RX Cummings N.J., Conerton I.F.;  
 RT "The *Bacillus subtilis* 168 chromosome from *sepe* to *kata*."  
 RL Microbiology 143:1855-1859 (1997).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=168;  
 RC MEDLINE=98044033; PubMed=9384377;  
 RX Kunst F., Ogasawara N., Moszer J., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertolo M.C., Bessieres P., Bolochin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Brann M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicof F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,  
 RA Fritsch C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Haisepf G., Guy B.J., Haga K., Haeuch U., Harwood C.R., Henaut A.,  
 RA Hillebert H., Holstegg S., Hosono S., Hullo M.F., Ilaya M., Jones L.,  
 RA Joriss B., Karamata D., Kasahara Y., Klaerr-Bianchand M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsreth G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarovic V.,  
 RA Lee S.M., Leillane A., Liu H., Maeda S., Muel C., Medigue C.,  
 RA Medina N., Melillo R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Paro T.M., Portetelle D., Portwollik S., Prescott A.M.,  
 RA Plessen E., Puig C., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
 RA Sekinuchi J., Sekowska A., Seror S.J., Serrot P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 RA Taseo V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viani A., Wambit R., Wedler H., Weitzenger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RA "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."  
 RT Nature 390:249-256 (1997).  
 RL Native 390:249-256 (1997).  
 CC -1- FUNCTION: REGULATES THE EXPRESSION OF EXTRACELLULAR-PROTEIN GENES  
 CC OF *BACILLUS SUBTILIS*.  
 CC -1- SIMILARITY: TO SEVERAL B. SUBTILIS RNA POLYMERASE SIGMA FACTORS.  
 CC -1- SIMILARITY: TO B.NATTO SENN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, M34826; AAA22750.1; -;  
 DR EMBL, 282044; CAB04806.1; -;  
 DR EMBL, 299108; CAB12709.1; ALT\_INIT.  
 DR PIR, B59705; B59705.  
 DR SDDbList; BG10747; sens.  
 DR Transcription regulation; DNA-binding; Complete proteome.  
 FW Transcription regulation; H-T-H MOTIF (BY SIMILARITY).  
 FT DNA BIND 11 31  
 FT SEQUENCE 65 AA; 7912 MW; 4A685B04179CE318 CRC64;

Thu Sep 9 07:45:55 2004

us-10-713-208-6\_copy\_1\_75.rsp

Page 8

```
Qy      20 MKDYMGKISKEKF-----LDIVVELEKLNWVAPQDLLEKCLKNHRIDLK---- 69
Db      3 VKKKGKRRRRKRYGNQILPPELLLEKNKREIL-NSAELMEELYM--KIDEKHTQC 57
Qy      70 -TKIOK 74
Db      58 VTKYKK 63
```

Search completed: September 9, 2004, 05:49:55  
Job time : 25 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:48:31 ; Search time 116 Seconds

(without alignments)  
203.999 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75

Perfect score: 373  
Sequence: 1 MAEIGEDLDKSDVSLIFLM.....LEKLNHRIDUKTKIOKY 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 99606

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phase:\*  
10: SP\_plant:\*  
11: SP\_todent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacterioph:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	59	15.8	72	2 09AHY8	09ahy8 bacteriobd
2	53	14.2	48	16 09KHE2	09keh2 bacillus ha
3	53	14.2	64	16 08R161	08r161 fusobacteri
4	53	14.2	66	16 08G306	08g306 brucella su
5	52	13.9	51	2 086256	086256 klebsiella
6	52	13.9	51	16 0813B1	0813b1 bacillus ce
7	51.5	13.8	45	2 09RSY9	09rsy9 thermotoga
8	50.5	13.5	66	16 08YER6	08yer6 bacillus me
9	50.5	13.5	68	16 031505	031505 bacillus su
10	50	13.4	75	5 Q7YWS2	Q7yws2 echinococcu
11	48.5	13.0	66	10 022226	022226 arabidopsis
12	48.5	13.0	75	5 0814Q7	0814q7 plasmodius
13	48	12.9	48	2 054872	054872 streptococc
14	47.5	12.7	50	11 0939P2	0939p2 cavia porce
15	47.5	12.7	61	2 093AT0	093at0 pseudomonas
16	47.5	12.7	62	9 064081	064081 bacterioph

17	47.5	12.7	62	16 031940	031940 bacillus su
18	47.5	12.7	69	5 09TV43	09tv43 caenorhabd
19	47	12.6	40	5 08TMD3	08tdm3 homo sapien
20	47	12.6	40	5 09NBX3	09nbx3 acropora pa
21	46.5	12.5	56	16 08RG68	08rg68 fusobacteri
22	46.5	12.5	62	16 07VGD6	07vgd6 helicobacte
23	46.5	12.5	71	16 089A09	089a09 clostridium
24	46.5	12.5	72	16 097TS9	097ts9 clostridium
25	46	12.3	45	16 09PMW4	09pmw4 campylobact
26	46	12.3	60	2 083291	083291 enterococcu
27	45.5	12.2	44	16 050707	050707 borrelia bu
28	45.5	12.2	58	16 099V64	099v64 staphylococ
29	45.5	12.2	60	9 064158	064158 bacterioph
30	45.5	12.2	60	16 031890	031890 bacillus su
31	45.5	12.2	67	6 08SP66	08sp66 macaca mula
32	45.5	12.2	69	16 08EGP6	08egp6 streptococc
33	45.5	12.2	69	16 08E191	08e191 streptococc
34	45.5	12.2	70	16 08RW99	08rw99 staphylococ
35	45.5	12.2	71	16 081BD5	081bd5 bacillus ce
36	45.5	12.2	75	10 08LC08	08lc08 arabidopsis
37	45	12.1	42	16 0830D9	0830d9 enterococcu
38	45	12.1	43	10 09SOE5	09soe5 zea mays (m
39	45	12.1	61	13 013142	013142 cyprinus ca
40	45	12.1	62	2 049059	049059 mycoplasma
41	45	12.1	62	16 08X579	08x579 escherichia
42	45	12.1	67	16 081T45	081t45 bacillus an
43	45	12.1	70	16 083BG3	083bg3 coxiella bu
44	45	12.1	73	16 07UYF1	07uyf1 rhodospirill
45	45	12.1	74	16 09KCA8	09kca8 bacillus ha

#### ALIGNMENTS

RESULT 1  
09AHY8 PRELIMINARY; PRT; 72 AA.  
ID 09AHY8  
AC 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Lrp (Fragment).  
GN Lrp.  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=29488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NC19;  
RX MEDLINE=21225535; PubMed=11325940;  
RA Cliche T.A., Bittum S.B., Horewilt A.R., Engsin J.C.;  
RT "A Phosphopantetheinyl Transferase Homolog is Essential for  
RT Photorhabdus luminescens To Support Growth and Reproduction of the  
RT Entomopathogenic Nematode Heterorhabditis bacteriophora.";  
RL J. Bacteriol. 183:3117-3126 (2001).  
CC -1 SIMILARITY: BELONGS TO THE ASNC FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
DR EMBL AF288086; AAK16099.1; -  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription; DNA-dependent; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR InterPro: IPR000485; HTH\_AsnC\_lrp.  
DR Pfam: PF01037; ASNC\_trans\_reg; 1.  
DR PRINTS: PRO0033; HTHASNC.  
DR SMART; SM00344; HTH\_ASN\_C; 1.  
DR PROSITE; PS00519; HTH\_ASN\_C\_FAMILY; 1.  
KW DNA-binding; Transcription; Transcription regulation.  
FT NON TER 72  
SQ SEQUENCE 72 AA; 8407 MW; 803332BD88EBD2DB CRC64;  
Query Match 15.8%; Score 59; DB 2; Length 72;

Best Local Similarity 31.0%; Pred. No. 1e+02;  
Matches 22; Conservative 11; Mismatches 20; Indels 18; Gaps 3;  
QY 5 GEDLDKSDVSLIFLKKDYMGKSKSEKSPFLDLVELEKTLNVAPODLLEKLNH 64  
DB 9 GKJDDRDRNINLELQD-----GRISN-----VELSKVGLSP-----TCLERVR 50  
QY 65 RIDLTKYQKY 75  
DB 51 RLROGRTNGY 61

## RESULT 2

Q9KEH2 PRELIMINARY; PRT; 48 AA.  
AC Q9KEH2; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein BH0860.  
GN BH0860.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RK MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AF001510; BAB04599.1; -  
DR PIR; H83759; H83759.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 48 AA; 5624 MW; CA22BE9AD436AC0 CRC64;

Query Match 14.2%; Score 53; DB 16; Length 48;  
Best Local Similarity 41.3%; Pred. No. 2.7e+02;  
Matches 19; Conservative 6; Mismatches 11; Indels 10; Gaps 2;

QY 24 MGRGKISKSEKSPFLDLVELEKTLNVAPODLLEKLNH 69  
DB 1 MGR-----KLSPFELVLENKELNLPDQLSKIEK-----RLDEX 36

## RESULT 3

Q8RI61 PRELIMINARY; PRT; 64 AA.  
AC Q8RI61; 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE This protein.  
GN FN1756.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;  
OC Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=2186394; PubMed=11899109;  
RA Khatatval V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fontein M., Kyprides N., Overbeek R.,  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).

DR EMBL; AE010479; AAL93871.1; -  
DR GO; G01006790; P:sulfur metabolism; IEA.  
DR InterPro; IPR003749; THIS.  
DR Pfam; PF02597; THIS; 1.  
KW Complete proteome.  
SQ SEQUENCE 64 AA; 7383 MW; 9B3E1E29B3902AF CRC64;

Query Match 14.2%; Score 53; DB 16; Length 64;  
Best Local Similarity 29.5%; Pred. No. 3.6e+02;  
Matches 20; Conservative 14; Mismatches 13; Indels 20; Gaps 4;

QY 1 MAE1-GEDLDKSDVSLIFLKK-----DYMGRKISKSEKSPFLDLVELEKTLNVA 49  
DB 1 MAE1GKYEELNDVLDLYLKNRYVRIVDY--NGDIVKKSDF-----EKINIKY 51  
QY 50 PQDLDL 56  
DB 52 TDKIRIV 58

## RESULT 4

Q8G3D6 PRELIMINARY; PRT; 66 AA.  
AC Q8G3D6; 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN BR0012.  
OS Brucella suis.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1330 / B1ovar 1;  
RK MEDLINE=22247741; PubMed=12271122.  
RA Paulsen I.T., Seshadri R., Nelson K.E., Eichen J.A., Heidelberg J.F.,  
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.V., Beaman M.O.,  
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
RA Nelson M.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,  
RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,  
RA Hoover D.L., Lindler L.E., Hailings S.M., Boyle S.M., Fraser C.M.;  
RT "The Brucella suis genome reveals fundamental similarities between  
RT animal and plant pathogens and symbionts.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
DR EMBL; AE014316; AAN28970.1; -  
DR TIGR; BR0012; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 66 AA; 7674 MW; FC12C02001F514AE CRC64;

Query Match 14.2%; Score 53; DB 16; Length 66;  
Best Local Similarity 48.0%; Pred. No. 3.7e+02;  
Matches 12; Conservative 5; Mismatches 4; Indels 4; Gaps 1;

QY 40 VELEKXNLVY---APDQDLLEKCL 60  
DB 24 IELEDDLVLSRIGVDDLDMDQEL 48

## RESULT 5

Q86256 PRELIMINARY; PRT; 51 AA.  
AC Q86256; 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Laucine responsive element (Fragment).  
GN LRP.  
OS Klebsiella oxytoca.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Klebsiella.  
OX NCBI\_TaxID=571;

```

RN [1]
RP SEQUENCE FROM N.A.
RA Albrecht C., Kleiner D.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ASNC FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: Y16963; CAA76565.1; -.
DR GO: GO:0005622; C:intracellular; IEA.
DR GO: GO:0003700; P:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR000485; HTH_AsnC_tlp.
DR PRINTS; PR00033; HTHASNC.
KM DNA-binding; Transcription; Transcription regulation.
FT NON_TER
SQ SEQUENCE 51 AA; 5857 MW; AE2B4C696F788F03 CRC64;

Query Match 13.9%; Score 52; DB 2; Length 51;
Best Local Similarity 32.8%; Pred. No. 3.6e+02;
Matches 20; Conservative 8; Mismatches 15; Indels 18; Gaps 3;

QY 5 GEDLDKDVSLIFLMKDYMGKGIKSKSKSFLLDVLVEKXNLVAPDQDLLEKCLNIH 64
DB 9 GKDLDIDNINLEKQD-----GRISN-----VELSKVGLSP-----TCLERVX 50

QY 65 R 65
DB 51 R 51

RESULT 6
Q813F1 PRELIMINARY; PRT; 51 AA.
AC Q813F1;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE Hypothetical cytosolic protein.
GN EC3144.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goldsman E., Larsen N., D'Souza M., Malinas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL: AE017008; AAP10086.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 51 AA; 5923 MW; C611DB65BD468 CRC64;

Query Match 13.9%; Score 52; DB 16; Length 51;
Best Local Similarity 39.4%; Pred. No. 3.6e+02;
Matches 13; Conservative 8; Mismatches 10; Indels 2; Gaps 2;

QY 10 KSDVSLIFLMKDYMGKGIKSKSKSFLLDVLV 40
DB 9 ESEISKAITHWEKDLGRGSIYKTDILRDIMI 41

RESULT 7
Q9RSY9 PRELIMINARY; PRT; 45 AA.
AC Q9RSY9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
```

```

DE RNA polymerase sigma factor (Fragment).
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93294870; PubMed=7685830;
RA Kim C.W., Markiewicz P., Lee J.J., Schierle C.F., Miller J.H.;
RT "Studies of the hyperthermophile Thermotoga maritima by random
RT sequencing of cDNA and genomic libraries. Identification and
RT J. Mol. Biol. 231:960-981(1993).
RL -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
CC HSPF: P00579; 1SIG.
DR GO: GO:0005899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0016987; F:sigma factor activity; IEA.
DR GO: GO:0003700; P:transcription factor activity; IEA.
DR GO: GO:0016740; P:transferase activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO: GO:0006352; P:transcription initiation; IEA.
DR InterPro: IPR009043; RNA_pol_sigma.
DR InterPro: IPR007627; Sigma70_r2.
DR Pfam: PF04542; sigma70_r2; 1.
DR PROSITE: PS00715; SIGMA70_1; 1.
KM DNA-binding; DNA-directed RNA polymerase; Sigma factor;
KM Transcription regulation; Transferase.
FT NON_TER
FT NON_TER
SQ SEQUENCE 45 AA; 5049 MW; F8ED52EB456B12F5 CRC64;

Query Match 13.8%; Score 51.5; DB 2; Length 45;
Best Local Similarity 38.9%; Pred. No. 3.6e+02;
Matches 14; Conservative 6; Mismatches 11; Indels 5; Gaps 1;

QY 6 EDDKSDVSLIFLMKDYMGKGIKSKSKSFLLDVLVE 41
DB 3 EXLITSNRLVLSIKRYMGSG-----LSFDLIDIE 33

RESULT 8
Q8YEE6 PRELIMINARY; PRT; 66 AA.
AC Q8YEE6;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical cytosolic protein EWE11932.
GN EWE11932.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujter C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykilds A., Reznik G.,
RA Tsvetkov L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elser P.H., Hagius S., O'Callaghan D., Leveson J.-C.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009626; AAL53113.1; -.
DR PIR: AF3493; AF3493.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 66 AA; 7658 MW; E6C9AD8C01F514AB CRC64;
```

KA LII A.;

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002510; AAB84347.1; -.
DR PIR: T00821; T00821.
SQ SEQUENCE 66 AA; 7424 MW; 689741BEC88A2D54 CRC64;
QY
Query Match 13.0%; Score 49.5; DB 10; Length 66;
Best Local Similarity 38.6%; Pred. No. 1.1e+03;
Matches 17; Conservative 9; Mismatches 13; Indels 5; Gaps 2;
DB 27 GKISKE---KSFELVLEL-KLNVAPDQDLLEKLNIRHID 65
10 GGSKKPNCGLKEFLSLKADLKTLSKEELLEERLKNVAR 53

RESULT 12
Q81407 PRELIMINARY; PRT; 75 AA.
AC Q81407;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN FFL2545C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Patti A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalom S.J., Sub B., Peterson J., Андлюи S.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.U., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Wengell C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrett B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL: AB014852; AAN36591.1; -.
KW Hypothetical protein.
SQ SEQUENCE 75 AA; 9211 MW; 79AAECF086CCDB69 CRC64;
QY
Query Match 13.0%; Score 48.5; DB 5; Length 75;
Best Local Similarity 31.6%; Pred. No. 1.2e+03;
Matches 18; Conservative 9; Mismatches 19; Indels 11; Gaps 3;
DB 26 RKISKEKSFELVLEKLNIVAPDQDLLEKLNIRHID-----LTKIKQY 75
7 RLVELQSDILSLIVLELDEN---DAQDLYDK-IKNYNNDDSEINRFRSMLQY 59

RESULT 13
Q54872 PRELIMINARY; PRT; 48 AA.
AC Q54872;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OC Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9008767; PubMed=2676973;
RA Prudhomme M., Martin B., Wejman V., Claverys J.P.;
RT "Nucleotide sequence of the Streptococcus pneumoniae hexB mismatch
RT repair gene: homology of hexB to MutL of Salmonella typhimurium and to

```

```

RT PMS1 of Saccharomyces cerevisiae."
RL J. Bacteriol. 171:5332-5338 (1989).
DR EMBL: M29686; AAB8599.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 48 AA; 5630 MW; 012929BDA3022B6CD CRC64;
QY
Query Match 12.9%; Score 48; DB 2; Length 48;
Best Local Similarity 33.3%; Pred. No. 8.7e+02;
Matches 11; Conservative 8; Mismatches 14; Indels 0; Gaps 0;
DB 41 ELEKLNIVAPDQDLLEKLNIRHIDLTKRIQ 73
12 EEKLNIVAPDQPTDKVQXVIEFAENLDREKIK 44

RESULT 14
Q99PT2 PRELIMINARY; PRT; 50 AA.
AC Q99PT2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Prostaglandin H synthase 1 (Fragment).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Ogura T., Asano K., Shiomi T., Fukunaga K., Nakamura M., Matsubara H.,
RA Lilly C.M., Drazen J.M., Yamaguchi K.;
RT "Cyclooxygenase-2 expression during allergic inflammation in guinea
RT pig lungs."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB054840; BAB21394.1; -.
DR HSP; P05979; IEQH.
DR GO: GO:0004601; F:peroxidase activity; IEA.
DR GO: GO:0006979; P:response to oxidative stress; IEA.
DR InterPro: IPR002007; Antim.peroxidase.
DR pfam: PF03098; An_peroxidase; 1.
DR PROSITE: PS50292; PEROXIDASE_3; 1.
FT NON TER
FT NON TER
SQ SEQUENCE 50 AA; 5759 MW; 1E7FCB0FFC4040D5 CRC64;
QY
Query Match 12.7%; Score 47.5; DB 11; Length 50;
Best Local Similarity 36.0%; Pred. No. 1e+03;
Matches 18; Conservative 8; Mismatches 15; Indels 9; Gaps 3;
DB 18 FLMKDYMRGRKISKEKSFELVLEKLNIVAPDQDLLEKLN 62
4 FGMRKYTSFGELTGDK---EWAALEEL-YGDIDALEFYGLLEKLPN 49

RESULT 15
Q93AT0 PRELIMINARY; PRT; 61 AA.
AC Q93AT0;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Sensor histidine protein kinase (Fragment).
GN CHEA.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B301D;

```

RA Wei X., Camborne E.D., El-Sheikh A.F., Takemoto J.Y., Klotz M.G.,  
RT "Regulation of flagellar motility in Pseudomonas syringae.",  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF417007; AL08572.1; -  
DR GO; GO:0016301; F:Kinase activity; IEA.  
DR GO; GO:0004871; F:Signal transducer activity; IEA.  
DR GO; GO:0007165; P:Signal transduction; IEA.  
DR InterPro; IPR008207; Hpt.  
DR InterPro; IPR008208; Hpt\_N.  
DR Pfam; PF01627; Hpt; 1.  
DR ProDom; PD003142; Hpt\_N; 1.  
DR PROSITE; PS50894; HPT; 1.  
KW Kinase.  
FT NON TER 61  
SQ SEQUENCE 61 AA; 6678 MW; E60A5B3F56A8FE5 CRC64;  
  
Query March 12.78; Score 47.5; DB 2; Length 61;  
Best Local Similarity 28.68; Pred.No.1.3e+03;  
Matches 14; Conservative 12; Mismatches 16; Indels 7; Gaps 3;  
  
Oy 19 LMKDYM-GRGKISKEKSFLLVVELEKLNIVAPDQLLEKCKIKNIHRI 66  
Db 9 ILQDFLVEAGEILLEQLS--EQLVLELS---RPDDADLNLAFRGFTIV 51

Search completed: September 9, 2004, 05:51:58  
Job time : 120 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:48:31 ; Search time 119 Seconds  
(without alignments)  
178.076 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75  
Perfect score: 373  
Sequence: 1 MABIGEDDKSDVSSLIFFLM.....LEKOLKNIRHIDUKTKIQKY 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 806123

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1:	geneeqp1980s:*
2:	geneeqp1990s:*
3:	geneeqp2000s:*
4:	geneeqp2001s:*
5:	geneeqp2002s:*
6:	geneeqp2003as:*
7:	geneeqp2003bs:*
8:	geneeqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	20.1	55	2	AAy33287 Human FLI
2	63	16.9	27	2	AAy33286 Viral DED
3	63	16.9	27	2	AAy33289 Viral DED
4	61.5	16.5	64	6	ABb99801 Amino acil
5	56	15.0	60	1	AAp71178 Bacillus
6	55	14.7	46	4	AAm16687 Peptide #
7	55	14.7	46	4	ABb35670 Peptide #
8	55	14.7	46	4	AAm29171 Peptide #
9	55	14.7	46	4	ABb30505 Peptide #
10	55	14.7	46	4	ABb21099 Protein #
11	55	14.7	46	4	AAm68864 Human bon
12	55	14.7	46	4	AAm56485 Human bra
13	55	14.7	46	4	ABg50524 Human liv
14	55	14.7	46	4	AAm04403 Peptide #
15	55	14.7	46	5	ABg38444 Human pep
16	54	14.5	49	4	AAm20621 Peptide #
17	54	14.5	49	4	ABb41825 Peptide #
18	54	14.5	49	4	AAm35621 Peptide #
19	54	14.5	49	4	ABb25545 Protein #
20	54	14.5	49	4	AAm75515 Human bon
21	54	14.5	49	4	AAm62697 Human bra
22	54	14.5	49	4	ABg57263 Human liv
23	54	14.5	49	5	ABg45071 Human pep
24	51.5	13.8	47	5	AAU09375 Thermotog
25	51	13.7	47	4	AAU09373 Staphyloc

26	49.5	13.3	62	5	ABp33224 Human ORF
27	49.5	13.3	71	4	AAm18411 Peptide #
28	49.5	13.3	71	4	ABb37453 Peptide #
29	49.5	13.3	71	4	AAm30888 Peptide #
30	49.5	13.3	71	4	ABb32201 Peptide #
31	49.5	13.3	71	4	ABb22746 Protein #
32	49.5	13.3	71	4	AAm70575 Human bon
33	49.5	13.3	71	4	AAm58128 Human bra
34	49.5	13.3	71	4	ABg52254 Human liv
35	49.5	13.3	71	4	AAm06010 Peptide #
36	49.5	13.3	71	5	ABg40238 Human pep
37	49	13.1	38	4	ABb37000 Peptide #
38	49	13.1	38	4	AAm70140 Human bon
39	49	13.1	39	4	ABg51850 Human liv
40	49	13.1	53	4	AAm17292 Peptide #
41	49	13.1	53	4	ABg51131 Human liv
42	49	13.1	53	4	AAm04977 Peptide #
43	49	13.1	68	2	AAw7599 Staphyloc
44	48.5	13.0	30	2	AAw45253 Peptide c
45	48.5	13.0	66	3	AAg60317 Arabidops

## ALIGNMENTS

```
RESULT 1
AAy33287
ID AAy33287 standard; peptide; 55 AA.
XX
AC AAy33287;
XX
DT 23-NOV-1999 (first entry)
XX
DE Human FLICE DED peptide fragment.
XX
KW DED: death effector domain; FLICE protein; screening; anti-HHV-6;
KW Viral envelope; apoptosis; multiple sclerosis; immunodeficient;
KW central nervous system; infection.
XX
OS Homo sapiens.
XX
PN DE19612182-A1.
XX
PD 23-SEP-1999.
XX
PF 19-MAR-1998; 98DE-01012182.
XX
PR 19-MAR-1998; 98DE-01012182.
XX
PA (FARB ) BAYER AG.
XX
PI Weber O, Hug H;
XX
DR WPI, 1999-528902/45.
XX
PT Identifying substances with anti-human herpes virus 6 activity useful for
PT treating multiple sclerosis and infections of the central nervous system.
XX
PS Disclosure; Page 3; 4pp; German.
XX
CC This invention describes a novel method to screen for effective anti-HHV-
CC 6 compounds which comprises incubating Pas-primed cells transfected with
CC human herpes virus 6 (HHV-6) viral envelope or fragments of the viral
CC envelope with putative anti-HHV-6 compounds and examining the apoptotic
CC behaviour of the cells. The method is used to identify new anti-human
CC herpes virus 6 compounds, targeted to a death effector domain (DED) -
CC homologous region in HHV. These are effective for treating multiple
CC sclerosis and infections of the central nervous system, especially in
CC immunodeficient patients. This sequence represents a human FLICE protein
CC DED region which is used to describe the method of the invention
XX
SQ Sequence 55 AA;
```





PA (INSP) INST PASTEUR.  
 PI Druille P, Gruener A;  
 XX MPI; 2003-129263/12.  
 XX DR N-PSDB; AB223335.  
 XX PT New polynucleotide from Plasmodium falciparum and derived protein, useful  
 PT as immunogen for antimalarial vaccines and for preparing diagnostic or  
 PT therapeutic antibodies.  
 XX PS Claim 5; Fig 1C; 115pp; French.  
 XX CC The present sequence is a Plasmodium falciparum antigen, designated  
 CC DG747. This antigen generates an interferon-gamma response. The protein  
 CC is localized on the surface of sporozoites and on the intra-hepatic stage  
 CC of parasites. The antigen, as well as nucleic acids encoding it, is  
 CC useful as immunogens/vaccines for protection against infection by  
 CC Plasmodium falciparum. They are useful in treating P. falciparum malaria  
 CC and for in vitro diagnosis of infection  
 XX SQ Sequence 64 AA;  
 OY Query Match 16.5%; Score 61.5; DB 6; Length 64;  
 DB Best Local Similarity 30.5%; Pred. No. 13;  
 Matches 18; Conservative 13; Mismatches 21; Indels 7; Gaps 2;  
 OY 18 FLMKYMGRGK---SKRSPFLDVLVELEKLNIVAPDQDLLEKCLNHRIDLKTKIQ 73  
 DB 2 FHMHDYIYDDRILYNDKERN---VTKSDNKNVILKSDYKCNKCNVILKSDNKNVILK 56  
 RESULT 5  
 ID AAP71178 standard; protein; 60 AA.  
 AC AAP71178;  
 XX 25-MAR-2003 (revised)  
 DT 10-MAR-2003 (revised)  
 DT 15-APR-1991 (first entry)  
 DE Bacillus exoprotease secretion promoting protein.  
 XX Exoprotease.  
 XX Bacillus sp.  
 XX OS JP61282400-A.  
 XX PD 12-DEC-1986.  
 XX PF 07-JUN-1985; 85GP-00123736.  
 XX PR 07-JUN-1985; 85GP-00123736.  
 XX PA (MITU) MITSUBISHI CHEM IND LTD.  
 XX WPI; 1987-025820/04.  
 XX DR N-PSDB; AAN71310.  
 XX PT Protein promoting extracellular prodn. of exoprotease - by use of  
 PT recombinant DNA technique.  
 XX PS Claim 3; Page 1060; 8pp; Japanese.  
 XX CC Protein may be produced from a bacterial expression vector, and has  
 CC extracellular exoprotease production promoting activity. (Updated on 10-  
 CC MAR-2003 to add missing OS field.) (Updated on 25-MAR-2003 to correct PA  
 CC field.)  
 XX SQ Sequence 60 AA;

Query Match 15.0%; Score 56; DB 1; Length 60;  
 Best Local Similarity 27.8%; Pred. No. 53;  
 Matches 15; Conservative 14; Mismatches 17; Indels 8; Gaps 1;  
 OY 28 KISKESFLDVLVELEKLNIVAPDQDLLEK-----CLKNHRIDLKTKIQ 73  
 DB 7 KLIDHKTIFIEYSPDLIELADIAGKGRPMEXVEYIEORCKNLIATIEOMIX 60  
 RESULT 6  
 ID AAM16687 standard; protein; 46 AA.  
 AC AAM16687;  
 DT 12-OCT-2001 (first entry)  
 DE Peptide #121 encoded by probe for measuring cervical gene expression.  
 XX Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX cervical cancer.  
 XX OS Homo sapiens.  
 XX PN WO200157278-A2.  
 PD 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US000670.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488901/53.  
 XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human cervical epithelial cells.  
 XX Claim 27; SEQ ID NO 21513; 487pp; English.  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 46 AA;  
 OY Query Match 14.7%; Score 55; DB 4; Length 46;  
 DB Best Local Similarity 30.6%; Pred. No. 51;  
 Matches 11; Conservative 12; Mismatches 13; Indels 0; Gaps 0;  
 OY 34 SFLDVLVELEKLNIVAPDQDLLEKCLNHRIDLK 69  
 DB 1 NLDIFIEWEKVIILGEGKDLIKVCAQINKSLK 36  
 RESULT 7  
 ABB35670

ID	ABB35670	standard; peptide; 46 AA.
XX		
AC	ABB35670;	
DT	04-FEB-2002	(first entry)
XX		
DE	Peptide #3176	encoded by human foetal liver single exon probe.
XX		
XX	Human; foetal liver;	gene expression; single exon nucleic acid probe.
OS	Homo sapiens.	
XX		
PN	MO2001:57277-A2.	
PD	09-AUG-2001.	
XX		
PF	30-JAN-2001; 2001MO-US000669.	
XX		
PR	04-FEB-2000; 2000US-0180312P.	
PR	26-MAY-2000; 2000US-0207456P.	
PR	30-JUN-2000; 2000US-00608408.	
PR	03-AUG-2000; 2000US-00633366.	
PR	21-SEP-2000; 2000US-0234687P.	
PR	27-SEP-2000; 2000US-0236359P.	
PR	04-OCT-2000; 2000GS-00024263.	
XX		
PA	(MOLE-) MOLECULAR DYNAMICS INC.	
XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-483447/52.	
XX		
PT	Human genome-derived single exon nucleic acid probes useful for analyzing	
XX	gene expression in human fetal liver.	
XX		
PS	Claim 27; SEQ ID NO 28305; 639pp + Sequence Listing; English.	
XX		
CC	The invention relates to a single exon nucleic acid probe for measuring	
CC	human gene expression in a sample derived from human foetal liver. The	
CC	single exon nucleic acid probes may be used for predicting, measuring and	
CC	displaying gene expression in samples derived from human fetal liver. The	
CC	present sequence is a peptide encoded by a single exon nucleic acid probe	
CC	of the invention. Note: The sequence data for this patent did not form	
CC	part of the printed specification, but was obtained in electronic format	
CC	directly from WIPO at <a href="http://wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a>	
XX		
SQ	Sequence 46 AA:	
Query Match	14.7%;	Score 55; DB 4; Length 46;
Best Local Similarity	30.6%;	Pred. No. 51;
Matches 11; Conservative 12; Mismatches 13; Indels 0; Gaps 0		
	34 SFLDLVLELEKLNVAPOQDLLEKCLNNHRIDIK 69	
	:  : : : : : : : : : : : : : : : :	
	1 NLLDRIEMEKRVILGEGKLDITIKVCAQINKSLIK 36	
Db		
RESULT 8		
ID	AAAM29171	
AC	AAAM29171	standard; protein; 46 AA.
XX		
XX	AAAM29171;	
XX		
DT	17-OCT-2001	(first entry)
XX		
XX	Peptide #3208	encoded by probe for measuring placental gene expression.
XX	Probe; microarray; human; placenta; antenatal diagnosis;	
KM	genetic disorder.	
XX		
XX	Homo sapiens.	
OS		
XX		
XX	WO200157272-A2.	

XX	09-AUG-2001.
PD	
PF	30-JAN-2001; 2001MO-US000663.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-488897/53.
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
XX	gene expression in human placenta.
PS	Claim 27; SEQ ID NO 29440; 654bp; English.
XX	
CC	The present invention relates to single exon nucleic acid probes (SENP;
CC	see AAI13135-AAI57546). The present sequence is a peptide encoded by one
CC	such probe. The probes are useful for producing a microarray for
CC	predicting, measuring and displaying gene expression in samples derived
CC	from human placenta. The probes are useful for antenatal diagnosis of
XX	human genetic disorders
SQ	Sequence 46 AA;
Query Match	14.7%; Score 55; DB 4; Length 46;
Best Local Similarity	30.6%; Pred. No. 51;
Matches	11; Conservative 12; Mismatches 13; Indels 0; Gaps 0;
DB	
QY	34 SPFLDVLVEIKINLVAPDQLDEKCKNTRIDLK 69 : :     :   :   :   :   :   1 NLDIFIEIKERKVIIGEGKDILRKVCQAQNKSLK 36
RESULT 9	
ABB30505	
ID	ABB30505 standard; peptide; 46 AA.
AC	ABB30505;
XX	
DT	01-FEB-2002 (first entry)
DE	Peptide #3156 encoded by breast cell single exon nucleic acid probe.
XX	
KW	Human; microarray; single exon probe; gene expression; breast; disease; cancer.
OS	Homo sapiens.
PN	WO200157271-A2.
PD	09-AUG-2001.
PF	30-JAN-2001; 2001MO-US000662.
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.
XX	





XX KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX KM tuberosus sclerotic; Gaucher's disease; Niemann-Pick disease;  
XX KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;  
XX KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX KM primary ciliary dyskinesia; pulmonary hypertension;  
XX KM hyaline membrane disease.  
XX Homo sapiens.  
OS  
XX  
XX NO2001086003-A2.  
PD  
XX  
XX 15-NOV-2001.  
PF  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00362386.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
P1  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
PS  
XX Claim 27; SEQ ID NO 28109; 6346p; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of probes  
XX; the novel set of probes which hybridise at high stringency to a nucleic  
XX acid expressed in the human lung; measuring gene expression in a sample  
XX derived from human lung, comprising (a) contacting the array with a  
XX collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of the  
XX array; identifying exons in a eukaryotic genome, comprising (a)  
XX algorithmically predicting at least one exon from genomic sequences of  
XX the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exon should be assigned to a single gene, a peptide comprising one  
XX of 12011 sequences, mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORF). The probes are used for gene expression  
XX analysis, and for identifying exons in a gene, particularly using human  
XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
XX tuberosus sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
XX Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
XX dyskinetic sequence is a peptide/protein encoded by a single exon probe of  
XX the invention. Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic format  
XX directly from WtPO at ftp.wipo.int/pub/published\_pct\_sequences

Thu Sep 9 07:45:52 2004

us-10-713-208-6\_copy\_1\_75.rag

Page 8

XX  
SQ Sequence 46 AA;

Query Match 14.7%; Score 55; DB 5; Length 46;  
Best Local Similarity 30.6%; Pred. No. 51;  
Matches 11; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Qy 34 SFLDUVVELEKTLVAPPDLDLEKCKLNIRIDLK 69  
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|  
Db 1 NLDIDFIEMEKRVILGEGKLDILKRVCAQINRSILK 36

Search completed: September 9, 2004, 05:54:04  
Job time : 124 secs

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:49:40 : Search time 814 Seconds  
(without alignments)  
29.548 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_1\_75  
Perfect score: 373  
Sequence: 1 MABGEDDKSDVSLIFLM.....LEKCKNIHRIIDKTKIQXY 75

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 467853

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09D\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	277	74.3	56	16 US-10-627-571-5	Sequence 5, Appl
2	277	74.3	56	16 US-10-627-571-7	Sequence 6, Appl
3	237	63.5	56	16 US-10-627-571-6	Sequence 6, Appl
4	225.5	60.5	59	16 US-10-627-571-4	Sequence 4, Appl
5	99	26.5	75	14 US-10-001-254-31	Sequence 31, Appl
6	90	24.1	75	9 US-09-852-768-66	Sequence 66, Appl
7	90	24.1	75	12 US-10-668-955-66	Sequence 66, Appl
8	79	21.2	56	16 US-10-627-571-9	Sequence 9, Appl
9	74	19.8	72	14 US-10-001-254-34	Sequence 34, Appl
10	68	18.2	56	16 US-10-627-571-8	Sequence 8, Appl
11	55	14.7	49	9 US-09-864-761-36397	Sequence 36397, A
12	54	14.5	49	9 US-09-864-761-40843	Sequence 40843, A
13	53	14.2	73	16 US-10-437-963-169508	Sequence 169508, A
14	52.5	14.1	66	16 US-10-627-571-3	Sequence 3, Appl
15	52	13.9	65	12 US-10-424-599-164893	Sequence 164893, A

16	52	13.9	70	12 US-10-424-599-244882	Sequence 244882, A
17	51.5	13.8	47	9 US-09-823-266-15	Sequence 15, Appl
18	51.5	13.8	47	14 US-10-155-419-15	Sequence 15, Appl
19	51	13.7	9	US-09-823-266-13	Sequence 13, Appl
20	51	13.7	47	14 US-10-155-419-13	Sequence 13, Appl
21	50	13.4	53	12 US-10-424-599-181787	Sequence 181787, A
22	50	13.4	52	12 US-10-335-977-8130	Sequence 8130, Ap
23	49.5	13.3	62	11 US-09-864-408A-4394	Sequence 4394, Ap
24	49.5	13.3	71	9 US-09-864-761-38044	Sequence 38044, A
25	48.5	13.3	75	16 US-10-767-701-50523	Sequence 50523, A
26	49	13.1	38	9 US-09-864-761-46700	Sequence 46700, A
27	49	13.1	53	9 US-09-864-761-46621	Sequence 46621, A
28	49	13.1	68	9 US-09-939-980-359	Sequence 359, App
29	49	13.1	71	16 US-10-437-963-183337	Sequence 183337, A
30	48.5	13.0	66	9 US-09-815-242-11591	Sequence 11591, A
31	48.5	13.0	68	16 US-10-767-701-55846	Sequence 55846, A
32	48	12.9	54	16 US-10-627-571-19	Sequence 19, Appl
33	48	12.9	72	16 US-10-437-963-168063	Sequence 168063, A
34	47.5	12.7	47	9 US-09-823-266-14	Sequence 14, Appl
35	47.5	12.7	47	14 US-10-155-419-14	Sequence 14, Appl
36	47.5	12.7	51	12 US-10-372-876-620	Sequence 620, App
37	47.5	12.7	51	14 US-10-097-065-620	Sequence 620, App
38	47.5	12.7	54	12 US-10-424-599-164109	Sequence 164109, A
39	47.5	12.7	63	11 US-09-864-408A-7010	Sequence 7010, Ap
40	47	12.6	54	14 US-10-429-386-28668	Sequence 28668, A
41	47	12.6	60	12 US-10-424-599-183136	Sequence 183136, A
42	47	12.6	71	12 US-10-424-599-271793	Sequence 271793, A
43	47	12.6	75	10 US-09-764-891-5023	Sequence 5023, Ap
44	47	12.6	75	14 US-10-205-428-435	Sequence 435, App
45	46.5	12.5	52	12 US-10-424-599-228956	Sequence 228956, A

# ALIGNMENTS

RESULT 1  
US-10-627-571-5  
Sequence 5, Application US/10627571  
Publication No. US20040082771A1  
GENERAL INFORMATION:  
APPLICANT: KASID, Usha N.  
APPLICANT: KUMAR, Deepak  
APPLICANT: GOKHALE, Prafulla  
APPLICANT: AHMAD, Imran  
TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-62 AND DIAGNOSTIC AND THERAPEUTIC USES  
FILE REFERENCE: 223316  
CURRENT APPLICATION NUMBER: US/10/627,571  
CURRENT FILING DATE: 2003-07-25  
PRIOR APPLICATION NUMBER: US 60/264,062  
PRIOR FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: PCT/US02/02212  
PRIOR FILING DATE: 2002-01-28  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 5  
LENGTH: 56  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
OTHER INFORMATION: CASH Alpha/Beta - fragment  
US-10-627-571-5

Query Match 74.3% Score 277, DB 16, Length 56,  
Best Local Similarity 100.0%, Pred. No. 2e-25,  
Matches 56, Conservative 0, Mismatches 0, Indels 0, Gaps 0;  
Oy 11 SVSSSLIFMKDYMGRGKISKEKSLDVLVELEKNIIVAPQDLLEKCKNIHRI 66  
Db 1 SVSSSLIFMKDYMGRGKISKEKSLDVLVELEKNIIVAPQDLLEKCKNIHRI 56

RESULT 2  
US-10-627-571-7

; Sequence 7, Application US/10627571  
; Publication No. US20040082771A1  
; GENERAL INFORMATION:  
; APPLICANT: KASID, Usha N.  
; APPLICANT: KUMAR, Deepak  
; APPLICANT: GOKHALE, Pratulla  
; APPLICANT: AHMAD, Imran  
; TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-S2 AND DIAGNOSTIC AND THERAPEUTIC USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 223316  
; CURRENT APPLICATION NUMBER: US/10/627,571  
; CURRENT FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: US 60/264,062  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: PCT/US02/02212  
; PRIOR FILING DATE: 2002-01-28  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: FLIP (L) - fragment  
US-10-627-571-7

Query Match 74.3%; Score 277; DB 16; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2e-25;  
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 11 SDVSSLIPLMKDYMGRGKISKEKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 66  
DB 1 SDVSSLIPLMKDYMGRGKISKEKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 56

RESULT 3  
US-10-627-571-6

; Sequence 6, Application US/10627571  
; Publication No. US20040082771A1  
; GENERAL INFORMATION:  
; APPLICANT: KASID, Usha N.  
; APPLICANT: KUMAR, Deepak  
; APPLICANT: GOKHALE, Pratulla  
; APPLICANT: AHMAD, Imran  
; TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-S2 AND DIAGNOSTIC AND THERAPEUTIC USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 223316  
; CURRENT APPLICATION NUMBER: US/10/627,571  
; CURRENT FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: US 60/264,062  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: PCT/US02/02212  
; PRIOR FILING DATE: 2002-01-28  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: FLIP (L) - fragment  
US-10-627-571-6

Query Match 63.5%; Score 237; DB 16; Length 56;  
Best Local Similarity 80.4%; Pred. No. 1.1e-20;  
Matches 45; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

11 SDVSSLIPLMKDYMGRGKISKEKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 66

DB 1 NDVSSLVFLTRDYTGKIAKDKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 56

RESULT 4  
US-10-627-571-4  
; Sequence 4, Application US/10627571  
; Publication No. US20040082771A1  
; GENERAL INFORMATION:  
; APPLICANT: KASID, Usha N.  
; APPLICANT: KUMAR, Deepak  
; APPLICANT: GOKHALE, Pratulla  
; APPLICANT: AHMAD, Imran  
; TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-S2 AND DIAGNOSTIC AND THERAPEUTIC USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: 223316  
; CURRENT APPLICATION NUMBER: US/10/627,571  
; CURRENT FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: US 60/264,062  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: PCT/US02/02212  
; PRIOR FILING DATE: 2002-01-28  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 59  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; OTHER INFORMATION: CASH Alpha/Beta - fragment  
US-10-627-571-4

Query Match 60.5%; Score 225.5; DB 16; Length 59;  
Best Local Similarity 76.3%; Pred. No. 2.7e-19;  
Matches 45; Conservative 8; Mismatches 3; Indels 3; Gaps 1;

CY 11 SDVSSLIPL--MKDYMGRGKISKEKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 66  
DB 1 NDVSSLVFLTRDYTGKIAKDKSFLLDIVELEKLNVAAPDQDLLEKCLKNHRI 59

RESULT 5  
US-10-001-254-31

; Sequence 31, Application US/10001254  
; Publication No. US20030049702A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, John C.  
; APPLICANT: Godzik, Adam  
; APPLICANT: Pawlowski, Krzysztof  
; APPLICANT: Fiorentino, Loredana  
; APPLICANT: Lee, Sung Hyung  
; APPLICANT: Roth, Wilfred  
; APPLICANT: Stemmer-Liwen, Frank  
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins  
; FILE REFERENCE: P-LJ 5037  
; CURRENT APPLICATION NUMBER: US/10/001,254  
; CURRENT FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: 60/301,889  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: 09/715,893  
; PRIOR FILING DATE: 2000-11-17  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Homo sapien

US-10-001-254-31

Query Match 26.5%; Score 99; DB 14; Length 75;  
Best Local Similarity 36.4%; Pred. No. 0.00039;  
Matches 24; Conservative 13; Mismatches 29; Indels 0; Gaps 0;



QY 3 EIGEDLDKSDVSLIFLMKDYMGKISKEKSFPLDVELEKLNVAPODLDLEKLN 62  
DB 9 DIGEQLDSEDLASLKFSLDTIPQKQEPIDALMLFORLOEKRMLESNLSFLKEILFR 68  
QY 63 IHRIDL 68  
DB 69 INRLDL 74

## RESULT 6

US-09-952-768-66  
Sequence 66, Application US/09952768  
Patent No. US20020035242A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
Fernandes-Alnemri, Teresa  
Litwack, Gerald  
Armstrong, Robert  
Tomaseilli, Kevin  
TITLE OF INVENTION: MCH4 AND MCH5, APOPTOTIC PROTEASE,  
NUCLEIC ACIDS ENCODING AND METHODS OF USE  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: Suite 6300, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/952,768  
FILING DATE: 10-Sep-2001  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Christiansen, William T.  
REGISTRATION NUMBER: 44,614  
REFERENCE/DOCKET NUMBER: 480140.424C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
FEATURES:  
NAME/KEY: Peptide  
LOCATION: 1..75  
OTHER INFORMATION: /note= "Mch5 A"  
SEQUENCE DESCRIPTION: SEQ ID NO: 66:  
US-09-952-768-66

Query Match 24.1%; Score 90; DB 9; Length 75;  
Best Local Similarity 34.8%; Pred. No. 0.0045;  
Matches 23; Conservative 13; Mismatches 30; Indels 0; Gaps 0;

QY 3 EIGEDLDKSDVSLIFLMKDYMGKISKEKSFPLDVELEKLNVAPODLDLEKLN 62  
DB 6 DIGEQLDSEDLASLKFSLDTIPQKQEPIDALMLFORLOEKRMLESNLSFLKEILFR 65  
QY 63 IHRIDL 68  
DB 66 INRLDL 71

## RESULT 7

US-10-668-955-66

Sequence 66, Application US/10668955  
Publication No. US20040054148A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
Fernandes-Alnemri, Teresa  
Litwack, Gerald  
Armstrong, Robert  
Tomaseilli, Kevin  
TITLE OF INVENTION: MCH4 AND MCH5, APOPTOTIC PROTEASE,  
NUCLEIC ACIDS ENCODING AND METHODS OF USE  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed Intellectual Property Law Group  
STREET: Suite 6300, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/668,955  
FILING DATE: 22-Sep-2003  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Laherty, Carol D.  
REGISTRATION NUMBER: 51,909  
REFERENCE/DOCKET NUMBER: 480140.424D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
FEATURES:  
NAME/KEY: Peptide  
LOCATION: 1..75  
OTHER INFORMATION: /note= "Mch5 A"  
SEQUENCE DESCRIPTION: SEQ ID NO: 66:  
US-10-668-955-66

Query Match 24.1%; Score 90; DB 12; Length 75;  
Best Local Similarity 34.8%; Pred. No. 0.0045;  
Matches 23; Conservative 13; Mismatches 30; Indels 0; Gaps 0;

QY 3 EIGEDLDKSDVSLIFLMKDYMGKISKEKSFPLDVELEKLNVAPODLDLEKLN 62  
DB 6 DIGEQLDSEDLASLKFSLDTIPQKQEPIDALMLFORLOEKRMLESNLSFLKEILFR 65  
QY 63 IHRIDL 68  
DB 66 INRLDL 71

## RESULT 8

US-10-627-571-9  
Sequence 9, Application US/10627571  
Publication No. US20040082771A1  
GENERAL INFORMATION:  
APPLICANT: KASID, Usha N.  
APPLICANT: KUNAR, Deepak  
APPLICANT: GOKHALE, Prafulla  
APPLICANT: AHMAD, Imran  
TITLE OF INVENTION: ANTI-APOPTOTIC GENE SCC-52 AND DIAGNOSTIC AND THERAPEUTIC USES  
FILE REFERENCE: 223316  
CURRENT APPLICATION NUMBER: US/10/627,571

```

; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/264,062
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/US02/02212
; PRIOR FILING DATE: 2002-01-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 9
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: FLICE (Casp8) - fragment
US-10-627-571-9
```

```

Query Match          21.2%; Score 79; DB 16; Length 56;
Best Local Similarity 25.5%; Pred. No. 0.064;
Matches 14; Conservative 21; Mismatches 20; Indels 0; Gaps 0;
```

```

QY      11 SDVSLFLMKDVMGRGKISKESFLDIYVELEKLNVAPODLDLEKLNHR 65
DB      1 SELSKFPLQSEISKCKDDNNLLDIFLEKRVILDEGKLDILKRVCAQINK 55
```

```

RESULT 9
US-10-001-254-34
; Sequence 34, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sung Hyung
; APPLICANT: Roth, Wilfried
; APPLICANT: Steiner-Blewen, Frank
; TITLE OF INVENTION: No. US20030049702A1a1 Death Domain Proteins
; FILE REFERENCE: P-1J 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-11-15
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 34
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-001-254-34
```

```

Query Match          19.8%; Score 74; DB 14; Length 72;
Best Local Similarity 29.5%; Pred. No. 0.34;
Matches 18; Conservative 12; Mismatches 27; Indels 4; Gaps 1;
```

```

QY      3 EIGEDLDKSVSLFLMKDVMGRGKISKESFLDIYVELEKLNVAPODLDLEKLN 62
DB      10 ELSEGDISENLKXWIFLKDSPK---TEMTSLFLAFLEKQKIDEDNLTCEDDLCKT 65
```

```

QY      63 I 63
DB      66 V 66
```

```

RESULT 10
US-10-627-571-8
; Sequence 8, Application US/10627571
; Publication No. US20040082771A1
; GENERAL INFORMATION:
; APPLICANT: KASID, Usha N.
; APPLICANT: KOMAR, Deepak
```

```

; APPLICANT: GOKALE, Prafulla
; APPLICANT: AHMAD, Imran
; TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-S2 AND DIAGNOSTIC AND THERAPEUTIC USES
; TITLE OF INVENTION: THEROOF
; FILE REFERENCE: 22316
; CURRENT APPLICATION NUMBER: US/10/627,571
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/264,062
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/US02/02212
; PRIOR FILING DATE: 2002-01-28
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 8
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: FLICE (Casp8) - fragment
US-10-627-571-8
```

```

Query Match          18.2%; Score 68; DB 16; Length 56;
Best Local Similarity 24.1%; Pred. No. 1.3;
Matches 13; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
```

```

QY      12 DVSLSFLMKDVMGRGKISKESFLDIYVELEKLNVAPODLDLEKLNHR 65
DB      2 ELRSKFLNNSEIPKCKLEDLSELEFVEKERTMLAENNTLKSICQVAK 55
```

```

RESULT 11
US-09-864-761-36397
; Sequence 36397, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DRIVEN SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36397
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007256.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
; OTHER INFORMATION: EST HUMAN HIT: BF095670.1, EVALUE 9.00e-19
; OTHER INFORMATION: SWISSPROT HIT: Q1490, EVALUE 7.00e-20
; US-09-864-761-36397

Query Match          14.7% Score 55; DB 9; Length 46;
Best Local Similarity 30.6%; Pred. No. 36;
Matches 11; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY      34 SFLDLVLEKLNVAPODLLKCKLNHRIDLK 69
DB      1 NLDLFIEMEKVILGEGKLDLKKVCAQINKSLK 36

RESULT 12
; Sequence 40843, Application US/09864761
; Patent No. US20020046763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40843
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121747.19
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.1
; OTHER INFORMATION: EST HUMAN HIT: BE395567.1, EVALUE 2.00e-23
; OTHER INFORMATION: SWISSPROT HIT: P19139, EVALUE 2.00e-24
; US-09-864-761-40843

Query Match          14.5% Score 54; DB 9; Length 49;
Best Local Similarity 37.5%; Pred. No. 51;
Matches 12; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY      40 VELEKLNVAPODLLKCKLNHRIDLKTK 71
DB      10 VHSNQHLYVSPALDFDLKLRVHQSRILTR 41

RESULT 13
; Sequence 169508, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169508
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67922C.1.pcp
; US-10-437-963-169508

Query Match          14.2% Score 53; DB 16; Length 73;
```

```

Best Local Similarity 37.9%; Pred. No. 1,18+02;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0.

QY 11 SDVSSLIPLMKDYMGRGKISKEKSFLLDY 39
   | : : : : : | : : : : |
Db 1 SNPSPLINLKTYIGANNIADQVERFDVLY 29

```

```

Query Match      13.9%   Score 52; DB 12; length 65;
Best Local Similarity 38.7%   Pred. No. 1.e+02;
Matches 12; Conservative 6; Mismatches 11; Indels 2; Gaps 1.

QY    36 LDVVELEKTNIVAPDQLDLKECLKRIHNI 66
       |::|::|||::|::|::|::|::|::|
Db     19 INIVLFEFKTKFFPDQTEFL--ISKILFRV 47

```

Search completed: September 9, 2004, 06:08:26  
Job time : 816 secs

```

RESULT 14
US-10-627-571-3
: Sequence 3, Application US/10627571
: Publication NO. US20040082771A1
: GENERAL INFORMATION:
: APPLICANT: KASID, Usha N.
: APPLICANT: KUMAR, Deepak
: APPLICANT: GOKHALE, Prafulla
: APPLICANT: AHMAD, Imran
: TITLE OF INVENTION: ANTI-APOTOPIC GENE SCC-S2 AND DIAGNOSTIC AND THERAPEUTIC USES
: TITLE OR INVENTION: THEREOF
: FILE REFERENCE: 223316
: CURRENT APPLICATION NUMBER: US/10/627,571
: CURRENT FILING DATE: 2003-07-25
: PRIOR APPLICATION NUMBER: US 60/264,062
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US02/02212
: PRIOR FILING DATE: 2002-01-28
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 3
: LENGTH: 66
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: MISC FEATURE
: OTHER INFORMATION: SCC-S2 - fragment
: US-10-627-571-3

```

Query Match	14.1%;	Score 52.5; DB 16;	Length 66;
Best Local Similarity	24.2%;	Pred. No. 1,1e+02;	
Matches	16;	Conservative 17;	Mismatches 28; Indels 5; Gaps 2;
Qy	6	EDLDSDVSESLIFLMDYMGGRGKISKE--KSFPLIVVEKEK---	LTVAAPDLEKCL 60
		: : :   : : :   : :   : :   : :   : :	
Db	1	DDTSSVEVDLRRVIRETYOKKGAENKIKLITIVYKAILRRNQPNQDELALNEKFK	60
		: : :   : : :   : :   : :   : :   : :	
Qy	61	KNIIRARI	66
		: : :   : : :   : :   : :   : :   : :	
Db	61	KKYVQOL	66

```

RESULT 15
US-10-424-599-164893
; Sequence 164893, Application US/10424599
; Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: Le Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
NUMBER OF SEQ ID NOS: 285684
SEO ID NO 164893
LENGTH: 65
TYPE: prt
ORGANISM: Glycine max
FEATURE:
;
; OTHER INFORMATION: Clone ID: PAT_MRT3847_119916C.1.dep
;S-10-424-599-164893

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: September 9, 2004, 05:54:54; Search time 17 Seconds

(without alignments)  
1001.524 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252

Sequence: 1 KQSVGAGTSYRNVLQAIAIQ.....YLAKPRMFFIQLNYVSDGQ 177

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 89042

Minimum DB seq length: 0  
Maximum DB seq length: 177

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.5	7.9	149	2 T43637	caspace protein 1C
2	70.5	7.6	135	2 AE2322	hypothetical prote
3	69	7.4	152	2 A97355	hypothetical prote
4	68.5	7.4	136	2 T08216	outer arm dynein 1
5	68	7.3	137	1 J04233	lysozyme (EC 3.2.1
6	67	7.2	125	2 A00297	probable membrane
7	65.5	7.0	164	2 T16321	hypothetical prote
8	64.5	6.9	89	2 E47758	retrovirus-related
9	64.5	6.9	114	2 T17929	hypothetical prote
10	64.5	6.9	174	2 H71267	conserved hypothe
11	63.5	6.8	167	2 S34363	hypothetical prote
12	63.5	6.8	171	2 A11060	probable acetyltra
13	63	6.8	172	2 E82053	probable fibrinoly
14	61.5	6.6	156	2 H86672	hypothetical prote
15	61.5	6.6	160	2 E82382	transcription regu
16	61	6.6	149	1 E69073	NADP-reducing hydr
17	61	6.6	160	2 H75062	flagella-related p
18	60.5	6.5	99	2 A70366	hypothetical prote
19	60	6.5	136	2 I51222	neurotrophin recep
20	60	6.5	157	2 S76052	hypothetical prote
21	59.5	6.4	154	2 F81917	hypothetical prote
22	59.5	6.4	156	2 T12893	hypothetical prote
23	58.5	6.3	120	2 E70479	dnak suppressor pr
24	58.5	6.3	146	2 AH2258	hypothetical prote
25	58.5	6.3	153	2 A87732	protein m1008.4 li
26	58.5	6.3	157	2 S35486	xy1.5 protein - pse
27	58.5	6.3	159	2 C97071	phosphoribosylcarb
28	58.5	6.3	165	2 AB2113	hypothetical prote
29	58	6.2	118	2 T14570	cytochrome b559 co

30	58	6.2	147	2 D69008	conserved hypothe
31	58	6.2	175	2 B97314	probable transcrip
32	57.5	6.2	160	1 E69186	conserved hypothe
33	57	6.1	104	2 A12257	hypothetical prote
34	57	6.1	119	2 T25062	hypothetical prote
35	57	6.1	162	2 T40004	hypothetical prote
36	57	6.1	173	2 T48097	glutathione peroxi
37	56.5	6.1	150	2 B23253	myosin A2 catalyti
38	56.5	6.1	161	2 A44821	general odorant-bl
39	56.5	6.1	173	2 T19101	hypothetical prote
40	56	6.0	94	2 T38652	hypothetical prote
41	56	6.0	95	1 A69697	ribosomal protein
42	56	6.0	104	2 G64340	hypothetical prote
43	56	6.0	109	2 D70077	hypothetical prote
44	56	6.0	154	2 H97860	hypothetical prote
45	55.5	6.0	148	2 C71353	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T43637  
caspace protein 1C - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C/Accession: T43637  
R.Shaham, S.  
U. Biol. Chem. 273, 35109-35117, 1998  
A/Title: Identification of multiple Caenorhabditis elegans caspases and their potential  
A/Reference number: 222587; MUID:99074291; PMID:9857046  
A/Accession: T43637  
A/Status: preliminary; translated from GB/EMBL/DBD  
A/Molecule type: mRNA  
A/Residues: 1-149 <SHA>  
A/Cross-references: EMBL:AF088287; NID:g4063371; PIDN:ACG98294.1; PID:g4063372  
C/Genetic: \*  
A/Note: csp-1

Query Match 7.9%; Score 73.5; DB 2; Length 149;  
Best local Similarity 28.7%; Pred. No. 9.3;  
Matches 27; Conservative 16; Mismatches 36; Indels 15; Gaps 4;

QY 55 YMAKSPKPGICILID-----CIGNETELARDT--FTSLGYEVOKFLHLSMGISQI 103  
DB 17 YEMNSMPRGTVLILSNENFKMERKRVGKDEVNLTFLFKLQYTVICKNLEASMLEA 76  
QY 104 IGGPACMPHEHRDYSFVCVLVSRG-GSOSVYGV 136  
DB 77 IKERF---EMARTDSILFLSHQDQASVFGID 107

##### RESULT 2

AE2322  
hypothetical protein alr4132 [imported] - Nostoc sp. (strain PCC 7120)  
C/Species: Nostoc sp. PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C/Accession: AE2322  
R.Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Watanabe, A.; Iriyuch  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata,  
DNA Res. 9, 205-213, 2001  
A/Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium An  
A/Reference number: AB1807; MUID:21595285; PMID:11759440  
A/Accession: AE2322  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-115 <KUR>  
A/Cross-references: GB:BA000019; PIDN:BA875831.1; PID:g17133267; GSPDB:GN00179  
A/Experimental source: strain PCC 7120  
C/Genetic: \*  
A/Note: alr4132  
C/Superfamily: Escherichia coli ybcQ protein



QY 121 -CVLVRG 127  
 DB 86 GCIASTG 93

RESULT 7  
 T16321  
 hypothetical protein F41C6.2 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C/Accession: T16321

R/Seisell, C.  
 submitted to the EMBL Data Library, October 1995  
 A/Description: The sequence of C. elegans cosmid F41C6.  
 A/Reference number: Z18495  
 A/Accession: T16321  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-164 <GET>  
 A/Cross-references: EMBL:U39745; NID:g1049470; PID:g1049471; PIDN:AAA80443.1; CESP:F41C6  
 C/Genetics:  
 A/Gene: CESP:F41C6.2  
 A/Introns: 51/3; 67/1

Query Match  
 Best Local Similarity 23.2%; Score 65.5; DB 2; Length 164;  
 Matches 32; Conservative 16; Mismatches 39; Indels 51; Gaps 6;

QY 61 PLGICHLI---DCIGN-----ETELRDFTSLGYEVCKLHSM 97  
 DB 7 PLRSCLTITYNDCDFVLRNSYVACVDVSRKKVDEPDWIKIKAPAGLISIMRLSS 66

QY 98 HGISQILGFCAMPFHRDYDSFVCL-----VSRGSGSVYGVDT-----HSGCLPLH 145  
 DB 67 VGLAQQYGVV-----EQAYKEVDVQVLEGAQVYEGEFTQVAVQVYVVALDHTGSPIS 121

QY 146 HIRRMFMDSCPYLAGXP 163  
 DB 122 N-----WIKGKP 128

RESULT 8  
 E47758  
 retrovirus-related reverse transcriptase homolog - Litiodendron chinense retrotransposon  
 C/Species: Litiodendron chinense  
 C/Date: 24-Feb-1994 #sequence\_revision 25-Apr-1997 #text\_change 05-Nov-1999  
 C/Accession: E47758

R/Voytas, D.F.; Cummings, M.P.; Konieczny, A.; Ausubel, F.M.; Rodermel, S.R.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 7124-7128, 1992  
 A/Title: copia-like retrotransposons are ubiquitous among plants.  
 A/Reference number: A46200; MUID:92357784; PMID:1379734

A/Contents: clone 2  
 A/Accession: E47758  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-89 <VOY>  
 A/Cross-references: GB:M94478; NID:g439435; PIDN:AAA33403.1; PID:g168308  
 A/Note: sequence extracted from NCBI backbone (NCBIP:111898)  
 C/Suprafamily: retrovirus-related polyprotein

Query Match  
 Best Local Similarity 26.8%; Score 64.5; DB 2; Length 89;  
 Matches 26; Conservative 13; Mismatches 23; Indels 35; Gaps 5;

QY 86 GREVCKF-----LHLSMGGISQILGFCAMPFHRDYDSFVCLVSRGSGSVYGVDT 139  
 DB 20 GTTVOGAENKVKLRSLIGLKLQKQV-----YKGFDSF---WMSRGFSKSEY----- 65

QY 140 SGLPLHHRMFMDSCPYLAG-KPKMFIFQNYVSD 175  
 DB 66 -----DYCVYFETINERKLISVFYVD 88

RESULT 9  
 T17929  
 hypothetical protein A426r - Chlorella virus pBCV-1  
 C/Species: Chlorella virus pBCV-1  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T17929

R/Graves, M.V.; Van Elteren, J.L.  
 submitted to the EMBL Data Library, May 1999  
 A/Reference number: Z18806  
 A/Accession: T17929  
 A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
 A/Residues: 1-114 <GRA>  
 A/Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96794.1  
 A/Experimental source: specific host Chlorella strain NC64A  
 C/Genetics:  
 A/Note: A426r

Query Match  
 Best Local Similarity 23.5%; Score 64.5; DB 2; Length 114;  
 Matches 23; Conservative 19; Mismatches 33; Indels 23; Gaps 3;

QY 4 VQAGTSYRNYLOAIOKSLDPNNFRPEPVKKSIQSEAPLPQSIPEERYKWKSKPLG 63  
 DB 1 MEGHQTYERMLKTVYRILE--ANSFRELIVPEKIKDTKKQIFQ----- 43

QY 64 ICLIIDICGNFTELLRDTFTSLG--YEVCKFLHSM 98  
 DB 44 ---IIDMICSNYNINKKFEQLNVDPEIONLOPLQNH 78

RESULT 10  
 H71267  
 conserved hypothetical protein TP0907 - syphilis spirochete  
 C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C/Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
 C/Accession: H71267

R/Praser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 reon, J.; Khailak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDi  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A/Reference number: A71250; MUID:9833270; PMID:9665876

A/Accession: H71267  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-174 <COU>  
 A/Cross-references: GB:AE001259; GB:AE000520; NID:g3323209; PIDN:AAC65859.1; PID:g33232  
 A/Experimental source: strain Nichols  
 C/Genetics:  
 A/Gene: TP0907

Query Match  
 Best Local Similarity 31.6%; Score 64.5; DB 2; Length 174;  
 Matches 18; Conservative 11; Mismatches 19; Indels 9; Gaps 3;

QY 120 VCVLRSRSGSVYGVDTGSG--LPLHHIRRMFMDSCPYLAGKPKMFIFQNYV 173  
 DB 123 VVLSTVEGGSSLVEVQRTHGVCYVPEFH--RFTID---VDVGRKKIHLRLQWIL 173

RESULT 11  
 S34363  
 hypothetical protein 18.3 - Salmonella typhimurium

C/Species: Salmonella typhimurium  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
 C/Accession: S34363

R/Persson, B.C.; Bjork, G.R.  
 submitted to the EMBL Data Library, June 1993  
 A/Description: Isolation of a gene (miaB) involved in hydroxylation of msz106A in tRNA

A/Reference number: S34359  
 A/Accession: S34363  
 A/Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-167 <PEER>  
A:Cross-references: EMBL:X73368; NID:g312705; PIDN:CA51784.1; PID:g312710

Query Match 6.8%; Score 63.5; DB 2; Length 167;  
Best Local Similarity 28.8%; Pred. No. 98;  
Matches 23; Conservative 17; Mismatches 25; Indels 15; Gaps 4;

QY 4 VGGAGTSSYRNVLQAAIQSLKDPNNFRPEPVKSIQSEAFIPQSIPEER--YKMKSK 60  
DB 99 IREGGLAKKLALMAL-----DHAREQGFRCVLETTAFLEALALYERLGFHEHSE 149  
QY 61 PLGICLIIDICIGNETELARD 80  
DB 150 PLGCTGHVDC--EVRMLKD 166

## RESULT 12

A11060  
Probable acetyltransferase STY4812 [imported] - Salmonella enterica subsp. enterica sero  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: This species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: A11060

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
Ch, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: A11060

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-171 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD06934.1; PID:g16505582; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY4812

Query Match 6.8%; Score 63.5; DB 2; Length 171;  
Best Local Similarity 28.8%; Pred. No. 1e+02;  
Matches 23; Conservative 17; Mismatches 25; Indels 15; Gaps 4;

QY 4 VGGAGTSSYRNVLQAAIQSLKDPNNFRPEPVKSIQSEAFIPQSIPEER--YKMKSK 60  
DB 103 IREGGLAKKLALMAL-----DHAREQGFRCVLETTAFLEALALYERLGFHEHSE 153  
QY 61 PLGICLIIDICIGNETELARD 80  
DB 154 PLGCTGHVDC--EVRMLKD 170

## RESULT 13

E82053  
Probable fimbrial assembly protein PilP VC2631 [imported] - Vibrio cholerae (strain N16  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 14-Jul-2003  
C:Accession: E82053

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.;  
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers,  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: E82053

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <HEI>  
A:Cross-references: GB:AE004339; GB:AE003852; NID:g9657211; PIDN:AA95772.1; GSPDB:GN001  
A:Experimental source: serogroup O1, strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC2631  
A:Map position: 1

C:Superfamily: Pilus assembly protein (PilP)

Query Match 6.8%; Score 63; DB 2; Length 172;  
Best Local Similarity 25.2%; Pred. No. 1.1e+02;  
Matches 32; Conservative 20; Mismatches 49; Indels 26; Gaps 7;

QY 57 MSKRLGICL-----IIDICIGNETELARDFTSTGYEVQKFLHSMHGISQILGGFACMP 111  
DB 1 MKNKPLGLMLSLVLVGCXANQ-----DDITSTVAQVERKARKVTKLPIL-EFEATP 54  
QY 112 --EHRDYSFVCVLSRSGSOSVYGVDP-----QTHSG-----PLHHR--MFGDSC 156  
DB 55 VQHKGRPFVLPKXALVQNPFLKACQWQVPAKXGPIERVYDHLRKGVSQGGSI 114  
QY 157 PYLAKP 163  
DB 115 SALVQTP 121

## RESULT 14

H86672  
Hypothetical protein yd1g [imported] - Lactococcus lactis subsp. lactis (strain IL140  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: H86672

R:Pollock, A.; Munker, P.; Manger, S.; Jallou, O.; Malarme, K.; Weissenbach, J.; Eh  
Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
A:Reference number: A86625; MUID:21235186; PMID:111337471  
A:Accession: H86672

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-156 <STO>  
A:Cross-references: GB:AE005176; PIDN:AAK04482.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yd1g

Query Match 6.6%; Score 61.5; DB 2; Length 156;  
Best Local Similarity 26.0%; Pred. No. 1.4e+02;  
Matches 25; Conservative 13; Mismatches 33; Indels 25; Gaps 3;

QY 19 IOKSLDPSNNFRPEPVKSIQES-----EAFIPQSIPEERYMKSKPIGICLIID 69  
DB 2 IDKKREDLTKLREMYINKRPESEKLDSDSKKFAFALSDDEKERTLNK----- 52  
QY 70 CIGNETELRDPTSTLGYEVQKFL--HLSMHGISQI 103  
DB 53 -----LELITDKLVTLDEKLDLAKNASADISEL 83

## RESULT 15

E82382  
Transcription regulator asnc family VCA1068 [imported] - Vibrio cholerae (strain N169  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: E82382

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers,  
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: E82382

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-160 <HEI>  
A:Cross-references: GB:AE004432; GB:AE003853; NID:g9658509; PIDN:AAF96962.1; GSPDB:GN  
A:Experimental source: serogroup O1, strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VCA1068  
A:Map position: 2  
C:Superfamily: regulatory protein asnc



Query Match 6.6%; Score 61.5; DB 2; Length 160;  
Best Local Similarity 25.0%; Pred. No. 1.4e+02;  
Matches 22; Conservative 12; Mismatches 33; Indels 21; Gaps 2;

QY 85 LGYEVOKPLHLSM-----HGISQLGQFACMPEHRDYDSPCYLVS----- 125

DB 61 LGDDVQVFIHRLDSSPSIFERRFAHAAVADIPEIACYSLSGDPDTMIKVRVDMKAYQA 120

QY 126 --RGGSSQSVYGVDTHTSGPLHHRMF 151

DB 121 FMGKLGSLPGVIGTRSEFVTEHKTSF 148

Search completed: September 9, 2004, 06:11:00  
Job time : 27 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 9, 2004, 05:52:06 ; Search time 12 Seconds  
(without alignments)  
768.035 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252

Perfect score: 930  
Sequence: 1 KQSVGAGTSYRNVLQAAIQ.....YLAGKPFQNFQNVSDGQ 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 38097

Minimum DB seq length: 0  
Maximum DB seq length: 177

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	7.7	176	YMO4_VIBPA	Q87mm7 vibrio para
2	68	7.3	137	LYC_BOMMO	P48816 bombyx mori
3	64.5	6.9	174	R1M4_TREPA	O83877 treponema p
4	64	6.9	101	VE7_HPV21	P50779 human papil
5	64	6.9	151	DKSA_BUCAP	Q8K95 buchiera ap
6	63.5	6.8	167	YUOM_SALT	Q08021 salmonella
7	60.5	6.5	99	Y754_AOUAE	Q66956 aquifex aeo
8	60	6.5	115	RL20_MYCPE	Q86UK7 mycoplasma
9	60	6.5	157	Y328_SANY3	Q85535 synechocyst
10	59.5	6.4	106	YF00_CLOTE	Q83446 clostridium
11	59	6.3	164	PBP1_EPIPO	Q95ve9 epiphyas po
12	58.5	6.3	157	YXG2_PSEPU	Q05092 pseudomonas
13	58	6.2	112	APGB_DROME	Q9rtu1 drosophila
14	57	6.1	161	YH81_XYLF	Q87AP2 xyliella fas
15	57	6.1	162	YB09_SCHPO	P87150 schizosacch
16	57	6.1	164	PBP2_EPIPO	Q95vF0 epiphyas po
17	57	6.1	173	GPX3_ARATH	Q91yb4 arabidopsis
18	56.5	6.1	113	VHED_BPS14	O21946 bacterioph
19	56.5	6.1	117	VHED_BPS10	O21945 bacterioph
20	56.5	6.1	161	OBP2_YANSE	P34119 manduca sex
21	56	6.0	95	Y1P3_SCHPO	Q9ut86 schizosacch
22	56	6.0	94	RL23_BACSU	P42924 bacillus su
23	56	6.0	104	Y327_METUA	O57773 methanococ
24	56	6.0	109	YX1H_BACSU	P42300 bacillus su
25	56	6.0	121	ENSA_BOVIN	Q97976 bos taurus
26	56	6.0	121	ENSA_MOUSE	Q97976 mus musculu
27	55.5	6.0	148	Y222_TREPA	Q83251 treponema p
28	55.5	6.0	174	VSH_TRTV	P33496 turkey thim
29	55	5.9	121	ENSA_HUMAN	O43768 homo sapien
30	55	5.9	122	RL18_LBPIP	Q92C20 leptospira
31	55	5.9	151	DKSA_BUCBP	Q89a23 buchiera ap
32	55	5.9	161	Y903_XYLFA	Q9pex5 xyliella fas
33	55	5.9	164	Y646_METUA	Q58062 methanococ

## ALIGNMENTS

RESULT 1	ID	YMO4_VIBPA	STANDARD;	PRT;	176 AA.
AC	Q87MM7				
DT	10-OCT-2003 (Rel. 42, Created)				
DT	10-OCT-2003 (Rel. 42, Last sequence update)				
DE	Hypothetical UPF015 protein VP2204.				
GN	VP2204				
OS	Vibrio parahaemolyticus				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;				
OC	Vibrionaceae; Vibrrio.				
OX	NCBI_TaxID=670;				
ON	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-RIMD 2210633 / Serotype O3:K6;				
RX	MEDLINE=22508454; PubMed=12620739;				
RA	Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,				
RA	Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,				
RA	Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;				
RT	Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism				
RT	distinct from that of V. cholerae.;				
RL	Lancet 361:743-749(2003)				
CC	-1- SIMILARITY: Belongs to the UPF015 family.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; AP005080; BAC60467.1; -				
DR	HAMAB; MF 01042; -; 1.				
DR	InterPro; IPR002625; Smr/MutS2.				
DR	Pfam; PF01713; Smr; 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 176 AA; 20200 MW; D6562D0E7D7EDC46 CRC64;				
Query Match	7.7%; Score 72; DB 1; Length 176;				
Best Local Similarity	23.9%; Pred. No. 7.5;				
Matches	37; Conservative 31; Mismatches 53; Indels 34; Gaps 9;				
QY	1 KQSVGAGTSYRNVLQAAIQSLKDPNNRREPEPKSIQF-----SFAFLPQSIPE 52				
DB	16 KENVQGV---KKLRQDITL---IQPFKNTKQKEIKRNRASSEFFYSDFEPLRNEE 68				
QY	53 -----ERYKKSKXPLGIC--LIIDCIG-NETSLRDTFTSLGYEVQKFLHLS-- 96				
DB	69 GPRRYARDVSTYEVKKLRGRGVYVDVLDHGMQGAKELEAMLAICYKNIHCAVC 128				
QY	97 MGIISQ-TIGQFA--CMEPHRDYDFVCVLVSRGG 128				
DB	129 OHGIGKHILKQKAPLMLAQHDPVMAFHQAPLEFGG 163				



```

RESULT 4
ID_ VET_HPV21 STANDARD; PRT; 101 AA.
AC PS0779;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE E7 protein.
GN E7.
OS Human papillomavirus type 21.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OX NCBI_TaxID=31548;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H.;
RL Submitted (09-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
CC ACTIVITIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U31779; AAA79395.1; -.
DR InterPro; IPR000148; Papv1_E7.
DR Pfam; PF00527; E7_1
KM Early protein; Transcription regulation; Oncogene;
KW DNA-binding; Trans-acting factor.
FT SITE 56 59 C-XX-C MOTIF-1.
FT SITE 89 92 C-XX-C MOTIF-2.
SQ SEQUENCE 101 AA; 11529 MW; B1D43A2A43C67B82 CRC64;

Query Match 6.9%; Score 64; DB 1; Length 101;
Best Local Similarity 30.5%; Pred. No. 23;
Matches 25; Conservative 13; Mismatches 34; Indels 10; Gaps 4;

QY 13 NVYQAIAIQKSLDPSNFRPEPVKKSIOSEAFIPQSIPBEYKMKSKPLGICLI-IDCI 71
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 16 NEDQPEVQ-----PVDLFCBELPSEGCETEHELPE---RTAYKYVT-PCGCCVKYKRI 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 GNETELLRDTFTSLGVEYOKFL 93
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 67 VNATQFAIRTFQNLFLFELQL 88
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID_ DSKA_BUCAP STANDARD; PRT; 151 AA.
AC O8K9U5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dnak suppressor protein homolog.
DN DSKA OR BUC8192.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canback B., Naeslund A.K., Eriksson A.-S.,
RA Werngren U.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RA "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -1- SIMILARITY: Contains 1 dksa/trar-type zinc finger.

```

```

CC -1- SIMILARITY: IN THE N-TERMINAL HALF, TO THE N-TERMINAL OF
CC B.SUBTILIS YTEA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB014095; AAM67757.1; -.
DR InterPro; IPR000962; Znf Deka/Trar.
DR Pfam; PF01258; ZF_dksa_trar, 1. FINGER; FALSE_NEG.
DR PROSITE; PS01102; DSKA_TRAR_ZN_FINGER.
KW Zinc-finger; Complete proteome.
FT ZN FING 114 138 TRAR/DSKA-TYPE.
FT ZN FING 114 138
SQ SEQUENCE 151 AA; 17720 MW; B6127A6769521EB9 CRC64;

Query Match 6.9%; Score 64; DB 1; Length 151;
Best Local Similarity 25.5%; Pred. No. 37;
Matches 24; Conservative 11; Mismatches 25; Indels 34; Gaps 4;

QY 10 SYRVYQAIAIQKSLDPSNFRPEPVKKSIOSEAFIP-----OSIPBEY 55
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 46 TWKQGLTEIHNTLLTYQDKSTNF-PDIPRAQGEFFSLRNDRSRKLIKIQETLK 104
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 KKKSKPLGIC-----LIIDC 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 105 KIKDKDFGVCNSCAVEIGIRLEARPTANLCIDC 138
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
ID_ YDGM_SALTY STANDARD; PRT; 167 AA.
AC Q08021;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical acetyltransferase yjgm (EC 2.3.1.-).
DN YDGM OR STM4473.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=LT2.
RC MEDLINE=94075212; PubMed=8253666;
RX Petersen B.C., Bjork G.R.;
RA "Isolation of the gene (mtaB) encoding the hydroxylase involved in
RA the synthesis of 2-methylthio-cis-ribozearin in rRNA of Salmonella
RA typhimurium and characterization of mutants.";
RL J. Bacteriol. 175:7776-7785(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewl N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston K., Wilson R.K.;
RA "Complete genome sequence of salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -1- SIMILARITY: Belongs to the acetyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X73368; CA51784.1; -.  
 DR EMBL; AE008910; AL223292.1; -.  
 DR PIR; S34363; S34363.  
 DR StyGene; SG10442; YJGM.  
 DR InterPro; IPR00182; GCSNacetyl\_trans.  
 DR Pfam; PF00583; Acetyltransf; 1.  
 KM Hypothetical protein; Transferase; Acyltransferase; Complete proteome.  
 SQ SEQUENCE 167 AA; 16329 MW; ACBF154FDB1CE87 CRC64;

Query Match 6.8%; Score 63.5; DB 1; Length 167;  
 Best Local Similarity 28.8%; Pred. No. 46;  
 Matches 23; Conservative 17; Mismatches 25; Indels 15; Gaps 4;  
 QY 4 VGGAGTGYRVVLAQAIQKSLKDPNNFREPVKSIQSEAFIPQSIPEER-YKMSK 60  
 DB 99 IRGGGLAKKALMAL-----DAREGFKCYLTETATLRALALYERLGEHISE 149  
 QY 61 PLGICLIIDICIGNETELRPD 80  
 DB 150 PLGGTGHVDC---EVRLKLD 166

RESULT 7  
 Y754\_AQUAE STANDARD; PRT; 99 AA.  
 ID Y754\_AQUAE  
 AC O66956;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AQ\_754.  
 GN AQ\_754.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "the complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus";  
 RT Nature 382:353-358(1998).  
 RL Nature 382:353-358(1998).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000706; AAC06922.1; -.  
 DR PIR; A70366; A70366.  
 KM Hypothetical protein; Coiled coil; Complete proteome.  
 FT DOMAIN 3 68  
 FT SEQUENCE 99 AA; 11582 MW; B03E17DD25D6B50 CRC64;

Query Match 6.5%; Score 60.5; DB 1; Length 99;  
 Best Local Similarity 32.0%; Pred. No. 49;  
 Matches 16; Conservative 13; Mismatches 20; Indels 1; Gaps 1;  
 QY 13 NVUQAIAQKSLKDPNNFREPVKSIQSEAFIPQSIPEER-YKMSK 62  
 DB 10 NLTKGALDEQSRABEYIRBEKVKAEITLELA-LETDIMEXELKLAKAL 58

RESULT 8  
 RL20\_MYCPE

ID RL20\_MYCPE STANDARD; PRT; 115 AA.  
 AC O8EUK7;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 50S ribosomal protein L20.  
 GN RPLT OR MYPE9150.  
 OS Mycoplasma penetrans.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=28227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HF-2;  
 RX MEDLINE=22354719; PubMed=12466555;  
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenti T., Furuya K.,  
 RA Yoshino C., Horino A., Shida T., Sasaki T., Hattori M.;  
 RT "The complete genomic sequence of Mycoplasma penetrans, an  
 RT intracellular bacterial pathogen in humans";  
 RL Nucleic Acids Res. 30:5293-5300(2002).  
 CC -1- FUNCTION: This protein binds directly to 23S ribosomal RNA and is  
 CC necessary for the in vitro assembly process of the 50S ribosomal  
 CC subunit. It is not involved in the protein synthesizing functions  
 CC of that subunit (By similarity).  
 CC -1- SIMILARITY: Belongs to the L20P family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AP004174; BAC44705.1; -.  
 DR HAMAP; MF\_00382; -; 1.  
 DR InterPro; IPR005813; Ribosomal\_L20.  
 DR InterPro; IPR005812; Ribosomal\_L20b/o.  
 DR Pfam; PF00453; Ribosomal\_L20; 1.  
 DR PRINTS; PR00062; RIBOSOMAL\_L20.  
 DR ProDom; PD002389; L20; 1.  
 DR TIGRFAMs; TIGR01032; rplT\_bact; 1.  
 DR PROSITE; PS00937; RIBOSOMAL\_L20; 1.  
 KM Ribosomal protein; rRNA-binding; Complete proteome.  
 SQ SEQUENCE 115 AA; 13796 MW; 1D7AB2BEB38B10A3 CRC64;

Query Match 6.5%; Score 60; DB 1; Length 115;  
 Best Local Similarity 23.0%; Pred. No. 65;  
 Matches 26; Conservative 17; Mismatches 36; Indels 34; Gaps 5;  
 QY 2 GSVQGA-----GTSYVNVQAAIQKS-----LKDPSNNFPE-----EPVKSIQSEAF 44  
 DB 18 KKAEGAKGINTSTYRNAQTYMAQSKYAYNDKRNKEDPFKWTARNAIRKENTYGA 77  
 QY 45 FLTPQSIPEERYKMSKPIGICLIIDICIGNETELIRDTFTSLGYEVQKFEHL 97  
 DB 78 FW-----HKLKKEIAL-----NRKMLSELAIONPSEFKFVSHV 113  
 RESULT 9  
 Y328\_SYNY3 STANDARD; PRT; 157 AA.  
 ID Y328\_SYNY3  
 AC Q55635;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative low molecular weight protein-tyrosine-phosphatase slr0328  
 DE (EC 3.1.3.48).  
 GN slr0328.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SIMILARITY: Belongs to the low molecular weight phosphotyrosine
CC protein phosphatase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63999; BAA10030.1; -.
DR PIR; S76052; S76052.
DR HSSP; S11064; 1D98.
DR InterPro; IPR000106; Low_mwt_PTPase.
DR Pfam; PF01451; LMWPC; 1.
DR PRINTS; PR00719; LMWPTPASE.
DR SMART; SM00226; LMWPC; 1.
DR Hypochemical protein; Hydrolase; Complete proteome.
KM ACT_SITE 7 7 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 12 12 BY SIMILARITY.
FT ACT_SITE 124 124 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 157 AA; 17470 MW; D642DB6D3A49751 CRC64;
SQ SEQUENCE
Query Match 6.5%; Score 60; DB 1; Length 157;
Best Local Similarity 27.9%; Pred. No. 94;
Matches 24; Conservative 14; Mismatches 26; Indels 22; Gaps 5;
QY 8 GTSRYNVLQALQKSLKDPNNRREPPYKKSIOESAPFLPQSIPEERYKMKSKPLG---- 63
DB 88 GDNVYRNIL-----AQDPAGGYHNK-VKMICDYTEKRGDRVDPYGGQA---GFEHV 136
QY 64 ICIIDIGENETELADTFSLGYEV 89
DB 137 IDLIEDACGN-----LITSLGKEL 155
RESULT 10
YF00 CLOTE STANDARD; PRT; 106 AA.
AC Q894NF;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypochemical UPF0145 protein CTC01500.
GN CTC01500.
OS Clostridium terani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Bruggemann H., Baumer S., Fricke W.F., Wieser A., Uessgang H.,
RA Decker I., Herzberg C., Martinek-Artas R., Weikl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium terani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -1- SIMILARITY: Belongs to the UPF0145 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015941; AAC06056.1; -.
DR HAMAP; MF_00338; -; 1.
DR InterPro; IPR002765; DUF74.
DR Pfam; PF01906; DUF74; 1.
DR Hypochemical protein; Complete proteome.
KM SEQUENCE 106 AA; 11422 MW; ECC1E69A5B0D1F7 CRC64;
SQ SEQUENCE
Query Match 6.4%; Score 59.5; DB 1; Length 106;
Best Local Similarity 25.0%; Pred. No. 66;
Matches 17; Conservative 16; Mismatches 34; Indels 1; Gaps 1;
QY 10 SYRYNVLQALQKSLKDPNNRREPPYKKSIOESAPFLPQSIPEERYKMKSKPLGICIID 69
DB 37 SFKNVVGEEI-KSYSEWVDVDTAVKMKVEAKNLDADAVINIRIAMTSMGSLTAVI 95
QY 70 CIGNETEL 77
DB 96 VSGTAVKV 103
RESULT 11
PBP1 EP1PO STANDARD; PRT; 164 AA.
ID PBP1 EP1PO
AC Q95VE9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Pheromone-binding protein 1 precursor (PBP 1).
GN PBP.
OS Epiphyas postvittana (light brown apple moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Tortricidae; Tortricidae; Tortricinae; Epiphyas.
OX NCBI_TaxID=65032;
RN [1]
RP SEQUENCE FROM N.A.; SEQUENCE OF 24-35, SUBUNIT, TISSUE SPECIFICITY,
RP AND POLYMORPHISM.
RC TISSUE=Antenna;
RX MEDLINE=22418600; PubMed=12530222;
RA Newcomb R.D., Sirey T.M., Rasseam M., Greenwood D.R.;
RT "Pheromone binding proteins of Epiphyas postvittana (Lepidoptera:
RT Tortricidae) are encoded at a single locus.";
RL Insect Biochem. Mol. Biol. 32:1543-1554(2002).
CC -1- FUNCTION: This major soluble protein in olfactory sensilla of male
CC moths serves to solubilize the extremely hydrophobic pheromone
CC molecules such as bombykol and to transport pheromone through the
CC aqueous lymph to receptors located on olfactory cilia (by
CC similarity).
CC -1- SUBUNIT: Monomer and disulfide-linked dimers.
CC -1- TISSUE SPECIFICITY: Antenna.
CC -1- POLYMORPHISM: 2 electrophoretic alleles are known: fast and slow.
CC The isoform shown here is the slow form.
CC -1- SIMILARITY: Belongs to the PBP/GOBP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF416588; AAL09027.1; -.
DR GO; GO:0005550; P:pheromone binding; TAS.
DR GO; GO:0006810; P:transport; NAS.
DR InterPro; IPR006072; Odorant.
DR InterPro; IPR006170; PBP_GOBP.

```

```
DR InterPro: IPR006625; PhbP.  
DR Pfam: PF01395; PhbP_GDBP. 1.  
DR PRINTS: PR00484; PhbP_GDBP.  
DR SMART: SM00708; PhbP. 1.  
KW Pheromone-binding; Pheromone response; Transport; Signal;  
KW Polymorphism.  
FT SIGNAL 1 23  
FT CHAIN 24 164 PHEROMONE-BINDING PROTEIN 1.  
FT DISULFID 42 77 BY SIMILARITY.  
FT DISULFID 73 130 BY SIMILARITY.  
FT DISULFID 119 139 BY SIMILARITY.  
SQ SEQUENCE 164 AA; 18362 MW; FDC633228469E2 CRC64;  
  
Query Match 6.3%; Score 59; DB 1; Length 164;  
Best Local Similarity 25.0%; Pred. No. 1.2e+02;  
Matches 26; Conservative 22; Mismatches 34; Indels 22; Gaps 7;  
  
OY 15 LQAALQKS---LKDPSNNPRE--EPYKSIQSEAFLEPOSIP-----EERYKKSKPL 62  
DB 17 LYQVBPQDVVKDMINFRKGLDCKKEIN-----LPDTINADFRFMDNDHVTNRDT 71  
OY 63 GICLIIDICGNTELLRDTSTLSGVEQKFLHSMHGSIQLIGQ 106  
DB 72 G-CAIM-CLASSKELVDTGLHGHNTLE---YAKOHGADETVAQ 110  
  
RESULT 12  
XYS2_PSEPU STANDARD; PRT; 157 AA.  
AC Q05092;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE XYDLEGF operon transcriptional activator 2.  
GN XYLS2.  
OS Pseudomonas putida.  
OC Plasmid TOL PDK1, and plasmid TOL pMW53.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H81 and MT53;  
RX MEDLINE=93065221; PubMed=1331988;  
RA Assinder S.V., de Marco P., Sayers J.R., Shaw L.E., Winsen M.K.,  
RA Williams P.A.;  
RT "Identical resolvases are encoded by Pseudomonas TOL plasmids pMW53  
RT and pDK1.";  
RL Nucleic Acids Res. 20:5476-5476(1992).  
CC -1- FUNCTION: REGULATORY PROTEIN OF THE TOL PLASMID XYL OPERON. XYLS  
CC ACTIVATES THE XYLXZLIEGHOQKIH OPERON REQUIRED FOR THE DEGRADATION  
CC OF TOLUENE, M-XYLENE AND P-XYLENE.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL  
CC REGULATORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: L02644; AAA71889.1; -  
CC EMBL: L02645; AAA71891.1; -  
CC PIR: S35486; S35486.  
CC InterPro: IPR000005; HTHARAC.  
CC Pfam: PF00165; HTH_ARAC. 2.  
CC PRINTS: PR00032; HTHARAC.  
CC SMART: SM00432; HTH_ARAC. 1.  
CC PROSITE: PS00041; HTH_ARAC_FAMILY_1; 1.  
CC PROSITE: PS01124; HTH_ARAC_FAMILY_2; 1.
```

```
KW Aromatic hydrocarbons catabolism; Transcription regulation; Activator;  
KW DNA-binding; Plasmid.  
FT DNA_BIND 55 74 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 157 AA; 17894 MW; A5FB035F3F8F674 CRC64;  
  
Query Match 6.3%; Score 58.5; DB 1; Length 157;  
Best Local Similarity 31.5%; Pred. No. 1.3e+02;  
Matches 29; Conservative 7; Mismatches 31; Indels 25; Gaps 5;  
  
OY 30 FREEPYKSIQ-ESBAFLPOSIPERRYKSKPLG-----ICLIIDICGNTELL 77  
DB 44 FIEENVKRSISLEQALALMSPRELYTMFEKHGTPTPMYIRNRKLECVRAKLSNPTTN 103  
OY 78 LRD-TSTLSGVEQKFLHSMHGSIQLIGQRA 108  
DB 104 IRTITEVALDY---GFLH-----LGRFA 123  
  
RESULT 13  
APGB_DROME STANDARD; PRT; 112 AA.  
AC Q9VTU1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Autophagy protein 12-like.  
GN CG10861.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.B., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaralides P.G., Scherer S.E., Li P.W., Hsieh R.A., Gale R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-U., Andrews-Piankovich C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bandaru D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.U., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostel C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ijzerman C.,  
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laslo P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,  
RA Palazzolo M., Pittman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,  
RA Reinert K., Remington K., Sanders R.D.C., Schaefer F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M.D., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";
```



```

RL Science 287;2185-2195(2000).
RN
RP REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield J., Bayraktarov L., Benman B.P.,
RA Betancourt B.R., Gelinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schröder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC
CC -1- FUNCTION: Required for autophagy (By similarity).
CC -1- SUBUNIT: Conjugated to autophagy protein 5-like (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: Conjugation of the G-112 to the K-132 of Autophagy protein 5-
CC like is a covalent modification that is essential for autophagy
CC (By similarity).
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE003542; AAF49955.2; ALT_SEQ.
CC Flybase; FBgn0036255; CG10661.
CC GO; GO:0005737; Cytoplasm; ISS.
CC GO; GO:0006914; P:autophagy; ISS.
CC InterPro; IPR007242; APG12.
CC Pfam; PF04110; APG12; 1.
CC DR Hypothetical protein; Autophagy; Ub1 conjugation pathway.
CC FT MOD_RES 112 112 CONJUGATION TO K-132 OF AUTOPHAGY PROTEIN
CC FT 5-LIKE (BY SIMILARITY).
CC SQ SEQUENCE 112 AA; 12865 MW; 755567D04D3DDAA4 CRC64;
Query Match 6.2%; Score 58; DB 1; Length 112;
Best Local Similarity 18.7%; Pred. No. 98;
Matches 26; Conservative 28; Mismatches 35; Indels 50; Gaps 7;
QY 50 IPERRYK-----KSKPLGICLIDICINETELRDFT-----SLGVEYKFL 93
DB 2 VPKVRIRKVFYFAFYLSQRLASVLCILNATGVPIIKRTWTDPNKTVGN-IQTFI 60
QY 94 H--LSNHG+SQILGQFACMPBEHRYDPSFVCVLSRGSQSYGVGDQTHSLPLHHRMF 151
DB 61 HKFLKIDASEQIF-----LY-VNQTAPAPFDIINKLV 92
QY 152 MEDSCPYLAGKPKMPTQN 170
DB 93 ---ECHGINKVLVYCKN 108
RESULT 14
YH81_XYLFT STANDARD; PRT; 161 AA.
AC 087AF2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein PD1781.
GN PD1781.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OC NCBI_TaxID=183190;

```

```

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Marino C.L., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferraz M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Abreu E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.B. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zartos L.G.,
RA Civerolo E.L., Simpson A.U.G., Almeida N.F. Jr., Setubal U.C.,
RA Kitajima U.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa.";
RL J. Bacteriol. 185:1018-1026(2003).
CC
CC -1- SIMILARITY: Belongs to the UPF0054 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE012559; AAC29615.1; -.
CC DR HAMAP; MF_00009; -.
CC DR InterPro; IPR002036; UPF0054.
CC DR Pfam; PF02130; UPF0054; 1.
CC DR PROSITE; PS01306; UPF0054; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 161 AA; 18013 MW; F690B38DB39CDB1 CRC64;
Query Match 6.1%; Score 57; DB 1; Length 161;
Best Local Similarity 20.6%; Pred. No. 1.9e+02;
Matches 26; Conservative 22; Mismatches 48; Indels 30; Gaps 5;
QY 7 AGTSRYNVQAIIQKSLCDP-----SNRFREPYKSIQSEATLPISIP 51
DB 23 AAVSRKVAATLQRIKADLAIRIVDEKGRALNHYRKQDVATVLPFPAQLPEFPF 82
QY 52 EERYVQKSKPLGICLIDICINETELRDFTSLGYEVO-KFLHLSMHGTSQILGQFACM 110
DB 83 -----KALKIPLDGIWCAPIAREA-TEGQKSLSHYATLHYGTHILGW---- 129
QY 111 PEHRYD 116
DB 130 -NREDH 134
RESULT 15
YB09_SCHPO STANDARD; PRT; 162 AA.
AC P87150;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C25H2.09 in chromosome II.
GN SPBC25H2.09.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4996;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:54:11 ; Search time 38 Seconds  
(without alignment)  
1469.651 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252  
Perfect score: 930  
Sequence: 1 KOSVQAGTSYRNVLQAAIQ.....YLAGKPKMFIONVYVSDGQ 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 372191

Minimum DB seq length: 0  
Maximum DB seq length: 177

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SP\_archaea:\*
  - 2: SP\_bacteria:\*
  - 3: SP\_fungi:\*
  - 4: SP\_human:\*
  - 5: SP\_invertebrate:\*
  - 6: SP\_mammal:\*
  - 7: SP\_mhc:\*
  - 8: SP\_organelle:\*
  - 9: SP\_phage:\*
  - 10: SP\_plant:\*
  - 11: SP\_rodent:\*
  - 12: SP\_virus:\*
  - 13: SP\_vertebrate:\*
  - 14: SP\_unclassified:\*
  - 15: SP\_virus:\*
  - 16: SP\_bacteriap:\*
  - 17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	277	29.8	78	6 Q8MJ18	Q8MJ18 macaca mula
2	94.5	10.2	104	4 Q9UG96	Q9UG96 homo sapien
3	78.5	8.4	131	11 Q9CXM4	Q9CXM4 mus musculu
4	73.5	7.9	149	5 Q9Y056	Q9Y056 caenorhabdi
5	70.5	7.6	135	16 Q8YPR0	Q8YPR0 anabaena sp
6	69	7.4	152	16 Q97CY2	Q97CY2 clostridium
7	68.5	7.4	136	10 Q04355	Q04355 chlamydomon
8	68	7.3	89	16 Q93J15	Q93J15 streptomyce
9	67	7.2	125	16 Q8ZDM3	Q8ZDM3 yersinia pe
10	66	7.1	96	10 Q9AXB5	Q9AXB5 oryza sativ
11	65.5	7.0	164	5 Q20267	Q20267 lixiorendro
12	64.5	6.9	89	10 Q06293	Q06293 paramectum
13	64.5	6.9	114	12 Q98478	Q98478 corytophane
14	64.5	6.9	125	13 Q9DDM5	Q9DDM5 lacerta viv
15	64.5	6.9	135	8 Q9B581	Q9B581 brachydanio
16	64.5	6.9	150	13 Q7SX13	Q7SX13

17	64	6.9	119	3 Q9P816	Q9P816 issatchenki
18	64	6.9	135	8 Q9B579	Q9B579 lacerta viv
19	63.5	6.8	135	8 Q9B0R7	Q9B0R7 lacerta viv
20	63.5	6.8	135	8 Q9B580	Q9B580 lacerta viv
21	63.5	6.8	135	8 Q9B586	Q9B586 lacerta viv
22	63.5	6.8	135	8 Q9B583	Q9B583 lacerta viv
23	63.5	6.8	135	8 Q9B0R8	Q9B0R8 lacerta viv
24	63.5	6.8	153	16 Q894A0	Q894A0 clostridium
25	63.5	6.8	171	16 Q82120	Q82120 salmonella
26	63	6.8	172	16 Q9KJ09	Q9KJ09 vibrio chol
27	62.5	6.7	110	10 Q84SM2	Q84SM2 oryza sativ
28	62.5	6.7	126	5 Q9NG09	Q9NG09 tribolium c
29	62.5	6.7	135	8 Q9B577	Q9B577 lacerta viv
30	62.5	6.7	135	8 Q9B1W8	Q9B1W8 lacerta viv
31	62.5	6.7	135	8 Q8LU47	Q8LU47 zootoca viv
32	62	6.7	101	12 Q84300	Q84300 human papil
33	62	6.7	155	10 Q81GP1	Q81GP1 arabidopsis
34	61.5	6.6	135	8 Q8M316	Q8M316 zootoca viv
35	61.5	6.6	135	8 Q9B582	Q9B582 lacerta viv
36	61.5	6.6	135	8 Q8M315	Q8M315 zootoca viv
37	61.5	6.6	135	8 Q9B578	Q9B578 lacerta viv
38	61.5	6.6	135	8 Q9MKK6	Q9MKK6 lacerta tri
39	61.5	6.6	139	8 Q8HH57	Q8HH57 podarcis ex
40	61.5	6.6	140	8 Q8HH56	Q8HH56 podarcis ex
41	61.5	6.6	141	8 Q8HH55	Q8HH55 podarcis ex
42	61.5	6.6	146	9 Q9B0B9	Q9B0B9 mycobacteri
43	61.5	6.6	156	16 Q9C1H6	Q9C1H6 lactococcus
44	61.5	6.6	160	16 Q9KXN4	Q9KXN4 vibrio chol
45	61.5	6.6	162	10 Q8S6B2	Q8S6B2 oryza sativ

ALIGNMENTS

RESULT 1	ID	Q8MJ18	PRELIMINARY;	PRT;	78 AA.
AC	Q8MJ18	01-OCT-2002 (TrEMBLrel. 22, Created)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)				
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Flic/caspase-1 inhibitory protein (Fragment).				
OS	Macaca mulatta (Rhesus macaque).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;				
OC	Cercopithecoidea; Macaca.				
OX	NCBI_TaxID=9544;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Petit F., Arnault D., Lelievre J.-D., Lecossier D., Hance A.J.,				
RA	Monceaux V., Ho Tsong Fang R., Hurtrel B., Amlisen J.-C.,				
RA	Estaque J.,				
RT	"Caspase-dependent and -independent cell death pathways characterize				
RT	pathogenic Simian Immunodeficiency Virus infection. Relationship with				
RT	disease evolution."				
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF530077; AAM95637.1; --				
FT	NON_TER	1			
FT	NON_TER	78			
FT	SEQUENCE	78 AA; 8971 MW; 6BD208095731A99 CRC64;			
QY	Query Match	29.8%; Score 277; DB 6; Length 78;			
QY	Best local similarity	73.1%; Pred. No. 9.8e-21;			
QY	Matches	57; Conservative 2; Mismatches 1; Indels 18; Gaps 1;			
DB	1 YKNUQAAIQSLKDPNNFR-----EPPVKSIOGESEAFIPQIPE 52				
QY	53 ERYKKSKPLGICLIIDC 70				
DB	61 ERYKKSKPLGICLIIDC 78				

## RESULT 2

09UG96 PRELIMINARY; PRT; 104 AA.  
 ID 09UG96  
 AC 09UG96;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein (fragment).  
 GN DKFZP586A181.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL050391; CAB366.2; -  
 DR HSSP: P29466; IICE  
 DR GO: GO:0030693; F:caspase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR001309; ICE\_p20.  
 DR InterPro: IPR002398; Peptidase\_C14.  
 DR Pfam: PF00656; Peptidase\_C14; 1.  
 DR SMART: SM00115; CASC; 1.  
 DR PROSITE: PS01122; CASPASE\_CYS; 1.  
 DR PROSITE: PS01121; CASPASE\_HIS; 1.  
 DR PROSITE: PS50208; CASPASE\_P20; 1.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 104 AA; 11587 MW; 10411DF1306C1432 CRC64;

Query Match 10.2%; Score 94.5; DB 4; Length 104;  
 Best Local Similarity 28.4%; Pred. No. 0.071;  
 Matches 27; Conservative 18; Mismatches 47; Indels 3; Gaps 2;

QY 78 LRDTFTSLGYEVOKFLHLSMHGISQILQFACMPFHRDYSFVCLVSRGSGOSVYGV-- 135  
 DB 3 MKELEGLDYSVVDVENLTLRDWESALRAFAFRPEKSSDSDFVLMSHGILGICGVH 62  
 QY 136 DQHTSGPLPHH-IRKMFAGDSCTYLAKGRMFTIQ 169  
 DB 63 DEKKPDLVLTDTFQIFNNENCLSLDKRKAIVIQ 97

## RESULT 3

09CKM4 PRELIMINARY; PRT; 131 AA.  
 ID 09CKM4  
 AC 09CKM4;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE 13 days embryo head cDNA, RIKEN full-length enriched library,  
 DE clone:3110059017, full insert sequence.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Toch M., Ishii Y.,  
 RA Aizawa T., Hara A., Fukunishi Y., Komu H., Adachi U., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochina H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schirral L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,  
 RA Brownstein M.U., Bulc C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli U., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK014231; BAB29219.1; -  
 DR HSSP: P42574; IGFV.  
 DR GO: GO:0030693; F:caspase activity; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR001309; ICE\_p20.  
 DR InterPro: IPR002398; Peptidase\_C14.  
 DR Pfam: PF00656; Peptidase\_C14; 1.  
 DR PRINTS: PR00376; ILBCEZYM.  
 DR PROSITE: PS50208; CASPASE\_P20; 1.  
 SQ SEQUENCE 131 AA; 14799 MW; 2B36D12AC47F62C7 CRC64;

Query Match 8.4%; Score 78.5; DB 11; Length 131;  
 Best Local Similarity 27.5%; Pred. No. 4.1;  
 Matches 30; Conservative 19; Mismatches 37; Indels 23; Gaps 6;

QY 28 NNPREBPV--KKSIOESAPFLPOSIPERRYCKSKPKICILID-----CIGN 73  
 DB 14 NNFEVKTNGSKSV--DSGIVLDS-----YKADYPMGICILINNNKPKHSTGMSRSGT 67  
 QY 74 ETEL--LRDTFTSLGYEVOKFLHLSMHGISQILQFACMP-EHRDYDSF 119  
 DB 68 DVDAANLRETFMGKLYQVRNNKNDLREDILFLMDSGRIEPIQSRFSFSG 116

## RESULT 4

09Y056 PRELIMINARY; PRT; 149 AA.  
 ID 09Y056  
 AC 09Y056;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Caspase-related protein 1C (Y483B.13c protein).  
 GN Y483B.13 OR CSP-1 OR Y483B.13c.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N2;  
 RX MEDLINE=99074291; PubMed=9857046;  
 RA Shahan S.;  
 RT "Identification of multiple Caenorhabditis elegans caspases and their  
 RT potential roles in proteolytic cascades."  
 RL J Biol. Chem. 273:35109-35117(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA McMurray A.A.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851316;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."  
 RT Science 282:2012-2018(1998).  
 DR EMBL: AF088287; AAC98294.1; -  
 DR EMBL: Z93393; CAD18880.1; -  
 DR PIR: T43637; T43637.  
 DR HSSP: P42574; 1CP3.  
 DR WormPeP: Y483B.13c; CB30017.  
 DR GO: GO:0030693; F:caspase activity; IEA.



Db 88 QVTVGQKGQAMRISRCLWDQT--NDNFASEYYENE-----SMYCVCQVYGLY 134

## RESULT 8

ID	AC	PRELIMINARY:	PRT:	89 AA.
CD	Q93J15:			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Hypothetical protein SC03993.			
DE	SC03993 OR SCBAC253.30C.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycinae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxId=1902;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2) / M145;			
RX	MEDLINE=1996410; PubMed=12000953;			
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,			
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,			
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			
RA	Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,			
RA	Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,			
RA	Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,			
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,			
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,			
RA	Hopwood D.A.;			
RT	"Complete genome sequence of the model actinomycete Streptomyces			
RT	coelicolor A3(2)." ;			
RL	Nature 417:141-147(2002) .			
DR	EMBL: AL539116; CAC44717.1; -			
KM	Hypothetical protein. Complete proteome.			
QW	SEQUENCE 89 AA; 10065 MW; 02FD9BB00B1068C2 CRC64;			

## RESULT 9

ID	ORGANISM	PRELIMINARY	PRT	125 AA
AC	OS2DW3			
AC	OS2DW3			
DT	01-MAR-2002 (TREMBLrel . 20, Created)			
DT	01-MAR-2002 (TREMBLrel . 20, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel . 24, Last annotation update)			
DE	Putative membrane protein.			
GN	Y02435.			
OS	Yersinia pestis.			
OS	Bacteria, Proteobacteria, Gammaproteobacteria, Enterobacteriales,			
OC	Enterobacteriaceae, Yersinia.			
OX	NCBI_TaxID=632;			
RP	[1] _SEQUENCE FROM N.A.			
RP	STRAIN=CO-92 / Biovar Orientalis;			
RC	MEDLINE=21470413 / PubMed=11586360;			
RA	Parhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,			
RA	Prentice M.B., Sobanina M., James K.D., Churcher C., Mungall K.L.,			
RA	Baker S., Basham D.Bentley S.D. Brooks K.Cedeno-Tarazaga A.M.,			
RA	Chillingworth T., Cronin A., Davies R.M., Davis B., Dougan G.,			
RA	Fellwell T., Hamlin N., Holtrop S., Jørgensen K., Kariyasek A.V.,			
RA	Leather S., Mølle S., Oyston P.C.F., Quail M., Rutherford S.G.,			
RA	Simmons W., Skelton T., Stevens T., Whitehead S., Garrett S.G.;			

RT "Genome sequence of Yersinia pestis, the causative agent of plague.",  
 RL Nature 413:523-527(2001).  
 DR EMBL; AJ41452; CAC91240.1; -.  
 DR PIR; AD0297; AD0297.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 125 AA; 14092 MW; 0C82AC4B2315289A CRC64;

## RESULT 10

ID	Q9AXB5	PRELIMINARY;	PRT,	96 AA.
AC	Q9AXB5			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2002	(TREMBLrel. 22, Last annotation update)		
DE	P0501G01.2 protein.			
GN	P0501G01.2			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehrhartoideae; Oryzoae; Oryza.			
OX	NCBI_TaxId=4530;			
RN	[1]_			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Nipponbare;			
RA	Sasaki T., Matsumoto T., Yamamoto K.;			
RT	"Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC			
RT	clone:P0501G01."			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF002819; BAB21073.1; -.			
DR	Gramene; Q9AXB5; -.			
Q	SEQUENCE 96 AA; 10793 MW; F53F7C4ABEEC640 CRC64;			

## RESULT 11

ID	Q20267	PRELIMINARY	PRT	164 AA.
AC	Q20267			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Hypochemical protein.			
GN	F1C6.2.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RX	MEDLINE=99069613; PubMed=9851916;			
RY	None;			
TY	Genome sequence of the nematode C. elegans. a platform for			

```
RT Investigating biology. The C. elegans Sequencing Consortium.";
RN Science 282:2012-2018 (1998).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Geisel C.;
RT "The sequence of C. elegans cosmid F41C6.";
DL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
DL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; U39745; AAA80443.1; -.
DR PIR; T16321; T16321.
DR WormRep; F41C6.2; CE04539.
KM Hypothetical protein.
SQ SEQUENCE 164 AA; 18790 MW; 6D73A7BD66BD88 CRC64;

Query Match
Best Local Similarity 7.0%; Score 65.5; DB 5; Length 164;
Matches 32; Conservative 16; Mismatches 39; Indels 51; Gaps 6;

QY 61 PLGICIL-----DGLGN-----ETELLRLFTSLGKRVCKLHLSM 97
DB 7 FPRSCFTITNCSDFVILRNSVVICVDSVRRKKVEPDVKIKAPAGLISIMRLSS 66
QY 98 HGISQILGQFACMPHRDYSEFVCL-----VSRGSGSVGVNQT-----HSGPLH 145
DB 67 VGLAQYGVV-----EQAVKEPDVQLBEVQAQYBEGFTVAVRVQVYRVVALDHTGSPIS 121
QY 146 HIRRMFMGDSPLYLAKKP 163
DB 122 N-----WILGKP 128

RESULT 12
ID Q06293 PRELIMINARY; PRT; 89 AA.
AC Q06293;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Reverse transcriptase (Fragment).
OS Liriodendron chinense (Chinese tulip tree).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; magnolids; Magnoliales; Magnoliaceae;
OC Liriodendron.
OC NCBI_TaxId=3414;
OX RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92357784; PubMed=1379734;
RT Voytas D.F., Cummings M.P., Konieczny A., Ausubel F.M., Rodermeil S.R.;
RT "Copia-like retrotransposons are ubiquitous among plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7124-7128 (1992).
DR EMBL; M94478; AAA33403.1; -.
DR PIR; E47758; E47758.
FT NON_TER 1 89
FT NON_TER 1 89
SQ SEQUENCE 89 AA; 10745 MW; 9279DDB25304AB65 CRC64;

Query Match
Best Local Similarity 6.9%; Score 64.5; DB 10; Length 89;
Matches 26; Conservative 13; Mismatches 23; Indels 35; Gaps 5;

QY 86 GRYVQCF-----LHLSHGISQILGQFACMPHRDYSEFVCLVSRGSGSVGVNQT 139
DB 20 GYVQAEKNCKVCKLRSLYGLKQILKQW-----YKFPDSF--WMSRFPSSKSEY----- 65
QY 140 SGLPLHIRRMFMGDSPLYLAG-KPMPFIQNYVVS 175
DB 66 -----DYCVTFETLNYEKPLISVFYDD 88
```

```
RESULT 13
ID Q98478 PRELIMINARY; PRT; 114 AA.
AC Q98478;
DT 01-FEB-1997 (TRENBLrel. 02, Created)
DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE A426R protein.
GN A426R.
OS Paramyxium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxId=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
positions 182 to 258.";
RL Virology 223:303-317 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Lisee A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homospemidine
synthase.";
RL Virology 263:254-262 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-L-3-Glucanase encoded by chlorella virus
PBCV-1.";
RL Virology 276:27-36 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gurnon J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42580; AAC96794.1; -.
DR PIR; T17925; T17925.
SQ SEQUENCE 114 AA; 13390 MW; 10425AED51BED1A5 CRC64;

Query Match
Best Local Similarity 6.9%; Score 64.5; DB 12; Length 114;
Matches 23; Conservative 19; Mismatches 33; Indels 23; Gaps 3;

QY 4 VQGAQTSYRNYLQAAIQSLKDPSPNPFREFPVKSGISEAFLPQSIPEERYKXKSKPLG 63
DB 1 MEGHQATVERMLKVLVYHLE--ANSFRELIVPEKIKDTKQIQP----- 43
QY 64 ICLITDCLGNTELLRLDFTSLG-----YEVQKFLHLSM 98
```

Db 44 ---IDMTCNMYNIMNDKDEQOLNVDPVEIOMLQPLONH 78

## RESULT 14

Q9DDMS PRELIMINARY; PRT; 125 AA.

AC Q9DDMS; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE C-mos (Fragment).  
 OS Corytophanes cristatus (Hemiteled basilisk).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Lepidodactylia; Squamata; Iguania; Iguanidae; Corytophaninae;  
 CC Corytophanes.  
 OC NCBI\_TaxID=144206;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Harris D.J.;  
 RT "Squamate relationships based on C-mos nuclear DNA sequences."  
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AF315390; AAC38568.1;  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004674; F:Protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006468; F:Protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KM ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 FT NON\_TER 1  
 FT NON\_TER 125  
 SQ SEQUENCE 125 AA; 13532 MW; 875B8801E2CE7204 CRC64;

Query Match 6.9%; Score 64.5; DB 13; Length 125;  
 Best Local Similarity 20.7%; Pred. No. 1e+02; 32; Indels 17; Gaps 5;  
 Matches 19; Conservative 24; Mismatches 32; Indels 17; Gaps 5;

QY 94 HLSMGGISQILGQFACMPHEDYDFVCYVSRGSGSQS---VYGV-----QTHSGPL 144  
 Db 20 HLNNNNVRYVAASTCAPDSQ--DSLGLIMEYVGNSTLHVIYGTCKTARRDDGLGC 77  
 QY 145 HHIRRMF--MDSCEYLAGKPMFFIONVV 173  
 Db 78 GHVSLSTTKALGYSCDIAG--LVFLSHLI 106

## RESULT 15

Q9B581 PRELIMINARY; PRT; 135 AA.

AC Q9B581; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Cytochrome b (Fragment).  
 GN CYTB.  
 OS Lacerta vivipara (Common lizard) (Zootoca vivipara).  
 CC Mitochondrion.  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Lepidodactylia; Squamata; Scleroglossa; Scincomorpha; Lacertoidae;  
 CC Lacertidae; Zootoca.  
 OC NCBI\_TaxID=8524;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21175762; PubMed=11277636;  
 RA Surget-Groba Y., Heulin B., Guillaume C.-P., Thorpe R.S.,  
 RA Kupriyanova L., Vogrin N., Maslak R., Mazotli S., Venczel M.,  
 RA Ghira I., Odierna G., Leontyeva O., Monney J.C., Smith N.;

RT "Intraspecific phylogeography of Lacerta vivipara and the evolution of  
 RT viviparity."  
 RL Mol. Phylogenet. Evol. 18:449-459(2001).  
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
 CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY). WHICH ARE NOT COVALENTLY  
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
 CC BOUND TO THE PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
 DR EMBL; AF247985; AAK29218.1;  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
 DR GO; GO:0005739; C:mitochondrion; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR005797; Cytb\_b6\_N.  
 DR Pfam; PF00033; Cytochrome\_B\_N; 1.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 KM Electron transport; Heme; Respiratory chain; Transmembrane; Transport;  
 KM Mitochondrion.  
 FT NON\_TER 135  
 FT NON\_TER 135  
 SQ SEQUENCE 135 AA; 15250 MW; B898FBA766770E66 CRC64;

Query Match 6.9%; Score 64.5; DB 8; Length 135;  
 Best Local Similarity 23.7%; Pred. No. 1.2e+02;  
 Matches 31; Conservative 12; Mismatches 51; Indels 37; Gaps 4;

QY 28 NNREPEPVKKSIOSEAFPLPSIPSEYRKMSKPLGICLIIDICGNTEHLLRDTFTSLGY 87  
 Db 4 MNRKQPLIKITINSFILLPFPNSISAMWNGSLGICLIILQITG----- 49  
 QY 88 EVQKFLHLSMGGISQILGQFACMPHEDYDFVCYVSRGSGSQS-----FVCVL--VSRGSG 129  
 Db 50 -----LFLAMHYTADISSAFSGVAHHRDVGHWLIRNLHANSAMFEICIVYHIGGLY 104  
 QY 130 QSVGVDOHHS 140  
 Db 105 YGSIVYTERWS 115

Search completed: September 9, 2004, 06:10:30  
 Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 05:50:00 ; Search time 54 Seconds

(without alignments)  
926.128 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252

Perfect score: 930  
Sequence: 1 KOSVGGAGTSTYRWVLAIAIQ.....YLAGKPKMFRTQWYVSDGQ 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 156107 seqs, 28547505 residues

Total number of hits satisfying chosen parameters: 1064358

Minimum DB seq length: 0  
Maximum DB seq length: 177

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	325	34.9	175	2	AAV05794
2	154	16.6	167	4	AAV31155
3	130	14.0	163	4	AAV68529
4	130	14.0	163	5	ABG38107
5	108	11.6	158	5	AAV16024
6	102	11.0	125	4	AAO02627
7	73	7.8	39	3	AB14254
8	71.5	7.7	101	6	ADA41128
9	71.5	7.7	101	6	ABR47936
10	71.5	7.7	101	6	ABR00176
11	71.5	7.7	101	7	ADB91676
12	71.5	7.7	101	7	ADC74344
13	71.5	7.7	102	2	AAV41346
14	67.5	7.3	133	2	ADB07674
15	66.5	7.2	102	5	ABP31818
16	66.5	7.2	169	5	ABU51964
17	65	7.0	100	6	ABP60619
18	65	7.0	117	2	AAV12157
19	64.5	6.9	174	6	ABU48709
20	64	6.9	130	6	ABM65939
21	64	6.9	139	4	AAU55771
22	63.5	6.9	139	6	ABM62290
23	63.5	6.8	166	6	ABU45215
24	63.5	6.8	171	6	ABU48266
25	63	6.8	113	6	ADB07666

25	63	6.8	134	6	ADB07666	ADB07666	Alloioiococ
27	63	6.8	153	5	AAE22127	AAE22127	Human 543
28	62.5	6.7	138	4	AAAB0170	AAAB0170	Coryneb
29	62.5	6.7	167	6	ABU26430	ABU26430	Aspergill
30	62	6.7	140	2	AAV35470	AAV35470	Chlamydia
31	62	6.7	161	6	ADA33022	ADA33022	Acinetoba
32	61.5	6.6	156	5	ABBS3691	ABBS3691	Lactococc
33	61	6.6	134	2	AAV55882	AAV55882	S. pneumo
34	61	6.6	142	6	AAE35761	AAE35761	Human SEC
35	61	6.6	159	5	ABP65784	ABP65784	Human pol
36	60.5	6.5	149	4	AAU27633	AAU27633	Human pro
37	60	6.5	109	4	AAV35040	AAV35040	Peptide #
38	60	6.5	109	4	ABB27326	ABB27326	Protein #
39	60	6.5	112	5	ABP31498	ABP31498	Human ORF
40	60	6.5	135	4	ABG62494	ABG62494	P. chryso
41	60	6.5	177	4	ABG09434	ABG09434	Novel hum
42	59.5	6.4	98	4	AAV94727	AAV94727	Human rep
43	59.5	6.4	129	5	AAE16023	AAE16023	Human cas
44	59.5	6.4	134	4	ABG11598	ABG11598	Novel hum
45	59.5	6.4	146	5	ABG63186	ABG63186	Bacillus

## ALIGNMENTS

RESULT 1  
AAV05794 standard; protein, 175 AA.

AAV05794; 02-AUG-1999 (first entry)

MRIT-D/S polypeptide.

MRIT-D/S; MACH related inducer of toxicity; human; apoptosis; anti-apoptotic; cancer; autoimmune disease; angiogenesis; atherosclerosis; neurodegenerative disease; Alzheimer's disease; Parkinson's disease; retinitis pigmentosa; stroke; AIDS; infection; aplastic anaemia; myocardial infarction; therapy; mutant.

XX Homo sapiens.  
OS Synthetic.  
PN W09918230-A2.  
PD 15-APR-1999.  
PF 07-OCT-1998; 98WO-US021132.  
PR 07-OCT-1997; 97US-00946226.  
PA (UNIV) UNIV WASHINGTON.  
PI Chaudhary PM;  
XX WFI, 1999-277275/23.  
PT Identifying regulators of MACH-related inducer of toxicity.  
PS Example 2; Page; 78pp; English.

XX The present sequence represents MRIT-D/S, comprising amino acid residues 306-880 of novel human MACH-related inducer of toxicity MRIT alpha 1 (see AAV05787). This deletion mutant was used to examine the interaction of MRIT alpha 1 with caspases. The Fricke-p20 domain interacted with MRIT alpha 1 but not with MRIT-D/S. The invention provides multiple isoforms of MRIT (see AAV05787-89), isolated active fragments of which have either pro-apoptotic or anti-apoptotic activity. Selective enhancers and inhibitors of MRIT apoptotic activity can be identified and used to treat diseases mediated by the dysfunction of programmed cell death or proliferation, such as cancer or a neurodegenerative disorder. Note: the present sequence is not shown in the specification but is derived from

CC the MRIT alpha 1 sequence given in figure 1F  
XX Sequence 175 AA;  
SQ

Query Match 34.9%; Score 325; DB 2; Length 175;  
Best Local Similarity 98.4%; Pred. No. 1.8e-30;  
Matches 60; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 117 DSFVCLVSRGSGSYGVQDTHSGLPFHHRMFWDSCPYLAGKPKMFIONVYVSDG 176  
DB 1 DSFVCLVSRGSGSYGVQDTHSGLPFHHRMFWDSCPYLAGKPKMFIONVYVSEG 60

QY 177 Q 177  
DB 61 Q 61

RESULT 2  
AAM31155  
ID AAM31155 standard; protein; 167 AA.  
XX  
AC AAM31155;  
XX  
DT 17-OCT-2001 (first entry)  
XX

DE Peptide #5192 encoded by probe for measuring placental gene expression.  
XX

KW Probe; microarray; human; placenta; antenatal diagnosis;  
KM genetic disorder.  
XX

CS Homo sapiens.  
XX

PN W0200157272-A2.  
XX

PD 09-AUG-2001.  
XX

PF 30-JAN-2001; 2001WO-US000663.  
XX

PR 04-FEB-2000; 2000US-0180312P.  
XX

PR 26-MAY-2000; 2000US-0207456P.  
XX

PR 30-JUN-2000; 2000US-00608408.  
XX

PR 03-AUG-2000; 2000US-00632366.  
XX

PR 21-SEP-2000; 2000US-0234687P.  
XX

PR 27-SEP-2000; 2000US-0236359P.  
XX

PR 04-OCT-2000; 2000GB-00024263.  
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX

DR WPI; 2001-488897/53.  
XX

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX

PS Claim 27; SEQ ID NO 31424; 654bp; English.  
XX

XX The present invention relates to single exon nucleic acid probes (SENP;  
CC see A131315-A157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX

SQ Sequence 167 AA;  
SQ

Query Match 16.6%; Score 154; DB 4; Length 167;  
Best Local Similarity 35.4%; Pred. No. 7.6e-10;  
Matches 35; Conservative 18; Mismatches 38; Indels 8; Gaps 3;

QY 78 LRDTFTSLGYEVQKFLHLSMHGISOILGQFACMPHRHYDSFVCLVSRGSGSYGVQD 137  
DB 2 LTTTFELHFEIKPHDCTVEQIYELIKIYQLM-DHSMNDFFICILISHGDKXGIYGTDS 60

QY 138 THSGPLHHRMFWDSCPYLAGKPKMFIO-----NY 171  
DB 61 QEA--FVETLSQFTGLKCPSLAGKPKMFIOACOGDNY 97

RESULT 3  
AAM65529  
ID AAM65529 standard; protein; 163 AA.  
XX

AC AAM65529;  
XX

DT 06-NOV-2001 (first entry)  
XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28835.  
XX

KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KM microarray; cancer; leukemia; lymphoma; myeloma.  
XX

OS Homo sapiens.  
XX

PN W0200157276-A2.  
XX

PD 09-AUG-2001.  
XX

PF 30-JAN-2001; 2001WO-US000668.  
XX

PR 04-FEB-2000; 2000US-0180312P.  
XX

PR 26-MAY-2000; 2000US-0207456P.  
XX

PR 30-JUN-2000; 2000US-00608408.  
XX

PR 03-AUG-2000; 2000US-00632366.  
XX

PR 21-SEP-2000; 2000US-0234687P.  
XX

PR 27-SEP-2000; 2000US-0236359P.  
XX

PR 04-OCT-2000; 2000GB-00024263.  
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX

DR WPI; 2001-488900/53.  
XX

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX

PS Example 4; SEQ ID NO 28835; 658bp + Sequence Listing; English.  
XX

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention  
XX

SQ Sequence 163 AA;  
SQ

Query Match 14.0%; Score 130; DB 4; Length 163;  
Best Local Similarity 33.3%; Pred. No. 5.8e-07;  
Matches 31; Conservative 17; Mismatches 43; Indels 2; Gaps 1;

QY 77 LRDTFTSLGYEVQKFLHLSMHGISOILGQFACMPHRHYDSFVCLVSRGSGSYGVQD 136  
DB 1 ILSHVQWLQFTVHNNVTVEEMVLOKQKCNPAADGDCVFCILTHGRGAVYSSD 60

QY 137 CTHSGPLHHRMFWDSCPYLAGKPKMFIO 169  
DB 61 E--ALIPREIMSHFTALQCPRLAKPKLFFIQ 91

RESULT 4  
ABG38107  
ID ABG38107 standard; peptide; 163 AA.  
XX

AC ABG38107;  
AC

XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 27772.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 PN MO200186003-A2.  
 XX  
 XX 15-NOV-2001.  
 PD  
 XX  
 PF 30-JAN-2001; 2001WO-US000665.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 XX  
 PS Claim 27; SEQ ID NO 27772; 634bp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridize at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridization of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridization to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary

CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Kargener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences  
 XX  
 XX Sequence 163 AA:  
 SQ  
 Query Match 14.0%; Score 130; DB 5; Length 163;  
 Best Local Similarity 33.3%; Pred. No. 5.8e-07;  
 Matches 31; Conservative 17; Mismatches 43; Indels 2; Gaps 1;  
 QY 77 LIRDFTSLGVEYOKFLHLSNHGISQILGFCACMPERHDYFVCULVGRGSGSYVCND 136  
 DB 1 ILSHVFQWLGFTVAIHNNVTKEWEMWVLOKCKNPAIADDCVFCLITGRGCAVYSSD 60  
 QY 137 QTHSGLPFHRIIRRMFGDSCPYLAGKRPFFIQ 169  
 DB 61 E-AIIPREIMSHFTALQCPRLAEXKXFFIQ 91  
 DE  
 XX  
 XX Human caspase-12 isoform, KM-F protein.  
 DE  
 XX  
 XX Human, cysteine-dependent aspartate-specific proteases; caspase-12; KM-F;  
 KW Parkinson's disease; ulcerative colitis; cytostatic; glomerulonephritis;  
 KW inflammatory bowel disease; hypersensitivity; rheumatoid arthritis; ALS;  
 KW amyotrophic lateral sclerosis; bronchitis; inflammatory cardiovascular;  
 KW neurodegenerative disease; Crohn's disease; Alzheimer's disease; cancer;  
 KW allergic rhinitis; cell proliferative disorder; asthma.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 40  
 FT /note= "Encoded by TC"  
 FT Misc-difference 63. .64  
 FT /note= "Encoded by ATGTGAGAT"  
 PN WO200185961-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 08-MAY-2001; 2001WO-US015103.  
 XX  
 PR 09-MAY-2000; 2000US-0203162P.  
 PA (PHAA) PHARMACIA & UPJOHN.  
 XX  
 PI Kletzien RF, Reardon IW, Welland KJ;  
 XX  
 PD WPI; 2002-082900/11.  
 DR N-PSDB; AAD26259.  
 XX  
 XX New human caspase-12 polynucleotides and polypeptides, useful for  
 PT screening modulators of caspase activity, e.g. inhibitors, especially for  
 PT treating e.g. inflammatory, cardiovascular or neurodegenerative diseases,  
 PT or cancer.  
 XX  
 XX Claim 1, Page 172; 207bp; English.  
 PS  
 XX The invention relates to purified, isolated caspase-12 polypeptides and  
 CC their polynucleotides. Cysteine-dependent aspartate-specific proteases  
 CC (caspases) are a family of proteases that cleave their substrates at

CC aspartic acid-X bonds. They are highly specific endopeptidases that  
 CC catalyse limited proteolysis. Caspase-12 polypeptides are useful for  
 CC screening modulators of caspase activity. Caspase12 inhibitors are useful  
 CC for preventing or treating disorders involving inappropriate apoptosis,  
 CC cardiovascular diseases, neurodegenerative disorders (Alzheimer's disease  
 CC Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis  
 CC (ALS)); inflammatory diseases (systemic inflammatory conditions and  
 CC conditions associated locally with migration and attraction of monocytes,  
 CC leukocytes and/or neutrophils); autoimmune diseases (rheumatoid or  
 CC inflammatory arthritis, acute glomerulonephritis, chronic glomerulonephritis,  
 CC inflammatory bowel diseases such as Crohn's disease, ulcerative colitis  
 CC and necrotising enterocolitis); allergic reactions (allergic asthma,  
 CC chronic bronchitis, acute and delayed hypersensitivity, allergic rhinitis  
 CC and cell proliferative disorder (cancer). Cancers that may be treated  
 CC comprise lymphomas, carcinomas or hormone-dependent tumours. The present  
 CC sequence is human caspase-12 isoform, KM-F protein  
 CC  
 SQ Sequence 158 AA;

Query Match 11.6%; Score 108; DB 5; Length 158;  
 Best Local Similarity 26.4%; Pred. No. 0.00025;  
 Matches 42; Conservative 19; Mismatches 70; Indels 28; Gaps 4;

QY 26 PSNNFREPVKSIQSEAFPLQSIPEERYKMKSPGLGICLTIDC-----IGNE 74  
 DB 11 PHAFHELTCKRA-----DEIYPVNEKERRTCTALNRNKEFNLYHNKNGSE 57  
 QY 75 TELLRDFTSLGYEVQKFLHLSMAGISQILGFCACMPHRDYDSFVCYLVSRSQSOSVY 133  
 DB 58 LDLLGMLLENLIGSVIKESLTAQEMETALRQFAHPEHOSDSTFLVFMSSHILNGIC 117  
 QY 134 GV--DQTHSGLPFLHTRFMFGDSCPYLAGKPKXPFPIQ 169  
 DB 118 GTKHWQDEPVDVLDHDTIFELFNNRNCSLKDKEKXVIMQ 156

RESULT 6  
 ID AA002627 standard; protein; 125 AA.  
 XX  
 AC AA002627;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 16519.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-00515126.

XX 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX DR N-PSDB; AAI82558.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 PT and treating e.g. leukemia, inflammation and immune disorders.

XX Claim 20; SEQ ID NO 16519; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibit activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 125 AA;

Query Match 11.0%; Score 102; DB 4; Length 125;  
 Best Local Similarity 24.4%; Pred. No. 0.00094;  
 Matches 32; Conservative 20; Mismatches 45; Indels 34; Gaps 4;

QY 17 AAIQSLKDPNNFREPVKSI-----QSEAFPLQSIPEERYKMKSPGLGICLTIDCIG 72  
 DB 6 ALYXRELFDPADKDXDHSRGIALVFYHRRFFVHLSPERR-----CTC 50  
 QY 73 NETELLRDFTSLGYEVQKFLHLSMAGISQILGFCACMPHRDYDSFVCYLVS 125  
 DB 51 SDRDNTRRFSDLGEGKCFNDLKAEBLLKIHVSTV-----HSDADCFVFLS 102  
 QY 126 RGSQSOSVYGVVD 136  
 DB 103 HGERHRYAYD 113

RESULT 7  
 ID AAB14254 standard; protein; 39 AA.  
 XX  
 AC AAB14254;

DT 09-FEB-2001 (first entry)

XX Partial mouse nedd-2 protein #1.

XX ced-3; virally induced cell death; apoptosis; gene therapy; neural;  
 KW muscular degenerative disease; myocardial infarction; stroke; aging;  
 KW interleukin-beta converting enzyme; ICE; mouse; Ice-ced 3 homologue;  
 KW Ich; nedd-2.

XX Mus sp.

XX Key Location/Qualifiers

XX FT Domain 28..32

XX US0683735-A.

XX 04-JUL-2000.

XX 10-JUN-1994; 94US-00258287.

XX 24-JUN-1993; 93US-00080850.

XX (GEMO) GEN HOSPITAL CORP.

XX Yuan J, Miura M;

XX WPI; 2000-464343/40.

XX DR N-PSDB; AAB72837.

XX New human Ich-1L and Ich-1S proteins for negative and positive regulation  
 PT of programmed cell death and for developing therapeutic methods for  
 PT diseases and conditions characterized by cell death, e.g. myocardial

PT infarction or stroke.  
 XX  
 PS Disclosure; Fig 9; 121p; English.  
 XX  
 CC The present sequence is a partial mouse nedd-2 protein. Three possible  
 CC reading frames were deduced for the coding sequence of the present  
 CC protein. The first reading frame encodes the present protein, which has a  
 CC potential QACG active domain. Nedd-2 is a member of a family of genes  
 CC involved in programmed cell death (apoptosis). Other family members  
 CC include: the ced-3 gene of *C. elegans* (AA072802), human interleukin-1beta  
 CC converting enzyme (ICE) (AAB14250), murine ICE1 (AAB14249), human Ich-1  
 CC and murine ICE2 (AAB14222). Ich-1 may play an important role in both the  
 CC positive and negative regulation of apoptosis. The Ich gene may be used  
 CC in gene therapy in disorders characterised by cell death e.g. neural and  
 CC muscular degenerative diseases, myocardial infarction, stroke, vitally  
 CC induced cell death and aging  
 XX  
 SQ Sequence 39 AA;  
 Query Match 7.8%; Score 73; DB 3; Length 39;  
 Best Local Similarity 53.8%; Pred. No. 0.55;  
 Matches 14; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 QY 144 LHHRRMEWGDSCEYLAGKKKFFIQ 169  
 Db 3 LQEVFRILPDNANCPSLQNKPKKFFIQ 28  
 RESULT 8  
 ADA41128  
 ID ADA41128 standard; protein; 101 AA.  
 XX  
 AC ADA41128;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human secreted protein.  
 XX  
 KW Human; secreted protein; cancer; hyperproliferative disorder;  
 KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;  
 KW anaemia; allergic reaction; asthma; cardiovascular disorder;  
 KW wound healing; cytostatic; immunosuppressive; neuroprotective;  
 KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;  
 KW vulnery; cardiant; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2002102993-A2.  
 XX  
 PD 27-DEC-2002.  
 XX  
 PF 19-MAR-2002; 2002WO-US008123.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-175238/17.  
 XX  
 PT New human secreted proteins and nucleic acid molecules, useful for  
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
 PT preventing or treating cancer or other hyperproliferative disorder,  
 PT asthma, allergies or AIDS.  
 XX  
 PS Claim 1; SEQ ID NO 1510; 3205bp; English.  
 XX  
 CC The invention relates to novel genes ADA39629-ADA40565 and proteins  
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,  
 CC treating or ameliorating medical conditions e.g. by protein or gene

CC therapy. The polypeptides, nucleic acid molecules, antibodies or their  
 CC fragments, and agonists or antagonists that bind to the polypeptide are  
 CC useful for preparing a diagnostic or pharmaceutical composition for  
 CC diagnosing or treating cancer or other hyperproliferative disorder. The  
 CC polypeptides and nucleic acid molecules are also useful for detecting,  
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer  
 CC or other hyperproliferative disorders including neoplasms, autoimmune  
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus  
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic  
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,  
 CC thrombocytopenia), allergic reactions including asthma or eczema,  
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory  
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders  
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,  
 CC fungal or viral infections including HIV/AIDS), or wound healing and  
 CC disorders of epithelial cell proliferation. The nucleic acids are also  
 CC useful for chromosome identification, radiation hybrid mapping or long-  
 CC range restriction mapping, as molecular weight markers, or as  
 CC hybridization or diagnostic probes. The polypeptides and antibodies are  
 CC useful for providing immunological probes for differential identification  
 CC of the tissues immunohistochemistry assays. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 101 AA;  
 Query Match 7.7%; Score 71.5; DB 6; Length 101;  
 Best Local Similarity 27.0%; Pred. No. 3.3;  
 Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;  
 QY 65 CLIDCIGNETELRLDPFTSLGCVQKFLHSMGIGSIQIGQFA-----CMEHR 114  
 Db 26 CLIFHCVSDSRSVNKEIV-----KVFHTSVGHSHSVQSHAFKFLVPPNAVPEQK 78  
 QY 115 DYD 117  
 Db 79 DPD 81  
 RESULT 9  
 ABR47936  
 ID ABR47936 standard; protein; 101 AA.  
 XX  
 AC ABR47936;  
 XX  
 DT 12-JUN-2003 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID 827.  
 XX  
 KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;  
 KW vulnery; antiinflammatory; neuroprotective;  
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200295010-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 19-MAR-2002; 2002WO-US009785.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-129429/12.  
 XX

PT Novel human secreted proteins, useful for detecting, preventing,  
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular  
PT disorders such as arrhythmia.  
XX  
XX Claim 13; SEQ ID NO 827; 1881bp; English.  
XX  
XX The present invention relates to novel human secreted proteins (ABR47633-  
CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins  
CC and their coding sequences are useful for the preparation of a diagnostic  
CC or pharmaceutical composition for diagnosing or treating a cardiovascular  
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary  
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, pulmonary disorders, renal disorders,  
CC proliferative disorders and/or cancerous diseases and conditions, for  
CC wound healing and epithelial cell proliferation, to treat inflammation or  
CC infection, for treating thrombosis and arteriosclerosis, for treating or  
CC preventing neural damage which occurs in neuronal disorders or  
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's  
CC disease, to enhance bone and periodontal regeneration and aid in tissue  
CC transplants or bone grafts, to prevent skin aging or hair loss, to  
CC stimulate growth and differentiation of haematopoietic cells and bone  
CC marrow cells when used in combination with other cytokines, to maintain  
CC organs before transplantation or for supporting cell culture of primary  
CC tissues, to increase or decrease differentiation or proliferation of  
CC embryonic stem cells, or to modulate mammalian characteristics or  
CC metabolism. Note: The sequence data for this patent was published in  
CC electronic format and is available from WIPO at  
CC [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 101 AA;  
XX  
Query Match 7.7%; Score 71.5; DB 6; Length 101;  
Best Local Similarity 27.0%; Pred. No. 3.3;  
Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;  
XX  
QY 65 CLIIDICGNETELLRDTFTSLGYEVQKFLHSMGSIQILGQFA-----CMEPHR 114  
Db 26 CIIFFCVSDRSRVNRETKV-----KPVHTSVHGVGHSFVQSAKFAFLVPEAVPEQK 78  
QY 115 DYP 117  
Db 79 DPD 81  
XX  
RESULT 10  
ABR00176  
ID ABR00176 standard; protein; 101 AA.  
XX  
AC ABR00176;  
XX  
DT 03-APR-2003 (first entry)  
XX  
DE Human gene 166 encoded secreted protein H1LEP53. SEQ ID NO:465.  
XX  
XX Human; secreted protein; digestive disorder; gastrointestinal disorder;  
KM mouth; oesophagus; stomach; small intestine; large intestine; liver;  
KM biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;  
KM immune disorder; inflammation; infection; wound healing; drug screening;  
KM chromosome identification; chromosome mapping; cytostatic;  
KM antiinflammatory; immunosuppressive; vulnereary; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200276488-A1.  
PN  
XX  
XX 03-OCT-2002.  
PD  
XX  
XX 19-MAR-2002; 2002WO-US008276.  
PF  
XX  
XX 21-MAR-2001; 2001US-0277340P.  
PR 19-JUL-2001; 2001US-0306171P.  
PR 13-NOV-2001; 2001US-0331287P.

XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX  
XX WPI; 2003-029900/02.  
DR  
XX  
XX N-PSDB; AB271355.  
DR  
XX  
XX New human secreted proteins and nucleic acids, useful for detecting,  
PT preventing, diagnosing, prognosticating, treating and/or ameliorating  
PT e.g. gastrointestinal diseases and disorders, or cancers.  
XX  
XX Claim 13; Page 1048-1049; 1216bp; English.  
XX  
XX AB271190-AB271478 represent cDNAs corresponding to 178 human secreted  
CC protein genes, and ABP00011-ABP00299 represent the proteins they encode.  
CC AB271479-AB271540 represent human secreted protein genomic fragments. The  
CC invention also encompasses antibodies specific for the secreted proteins,  
CC the use of the secreted proteins in drug screening, and recombinant  
CC vectors and host cells comprising a nucleic acid of the invention. The  
CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
CC fragments specific for the secreted proteins, and modulators of protein  
CC activity are useful for diagnosing, treating, ameliorating or preventing  
CC digestive disorders. Such conditions include disorders of the mouth,  
CC oesophagus, stomach, small intestine, large intestine, liver, biliary  
CC tract and pancreas, and include cancers of these organs and tissues. The  
CC secreted proteins and their nucleic acids may also be used in the  
CC treatment of immune disorders, inflammation, infection,  
CC hyperproliferative disorders, and to promote wound healing. Nucleic acids  
CC of the invention may be used for chromosome identification, chromosome  
CC mapping, in gene therapy, for identifying individuals from minute  
CC biological samples, as hybridisation probes, and as molecular weight  
CC markers. The present sequence represents a human secreted protein of the  
CC invention  
XX  
SQ Sequence 101 AA;  
XX  
Query Match 7.7%; Score 71.5; DB 6; Length 101;  
Best Local Similarity 27.0%; Pred. No. 3.3;  
Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;  
XX  
QY 65 CLIIDICGNETELLRDTFTSLGYEVQKFLHSMGSIQILGQFA-----CMEPHR 114  
Db 26 CIIFFCVSDRSRVNRETKV-----KPVHTSVHGVGHSFVQSAKFAFLVPEAVPEQK 78  
QY 115 DYP 117  
Db 79 DPD 81  
XX  
RESULT 11  
ADB91676  
ID ADB91676 standard; protein; 101 AA.  
XX  
AC ADB91676;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human secreted protein #SEQ ID 622.  
XX  
XX Secreted protein; gene therapy; antidiabetic; diabetes; human.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003004622-A2.  
PN  
XX  
XX 16-JAN-2003.  
PD  
XX  
XX 19-MAR-2002; 2002WO-US008124.  
PF  
XX  
XX 21-MAR-2001; 2001US-0277340P.  
PR 19-JUL-2001; 2001US-0306171P.  
PR 13-NOV-2001; 2001US-0331287P.

```
XX (HUMA-) HUMAN GENOME SCI INC.
PA 19-JUL-2001; 2001US-0277340P.
XX 13-NOV-2001; 2001US-0306171P.
XX 13-NOV-2001; 2001US-0331267P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX MPI; 2003-229407/22.
XX Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 622; 1537pp; English.
XX
XX The invention relates to isolated nucleic acid molecules ADB91065-
XX ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
XX ADB91834. Also disclosed is a recombinant vector comprising a
XX polynucleotide of the invention, and a recombinant host cell comprising
XX the recombinant vector. The polypeptide of the invention is useful in
XX identifying a binding partner by contacting the polypeptide with a
XX binding partner, and determining whether the binding partner increases or
XX decreases activity of the polypeptide. The polypeptide, polynucleotide,
XX antibody or its fragment, agonist or antagonist are useful for preparing
XX a pharmaceutical composition for diagnosing or treating diabetes or
XX conditions related to diabetes. The present sequence is that of the human
XX immunoglobulin Fc portion used to generate fusion proteins, increasing
XX the stability of the fused protein as compared to the secreted protein
XX only. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcc_sequences.
XX
SQ Sequence 101 AA;
XX
XX Query Match 7.7%; Score 71.5; DB 7; Length 101;
XX Best Local Similarity 27.0%; Pred. No. 3.3;
XX Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;
XX
QY 65 CLIIDIGNETELRLDPTSLGYEVQKFLHSMGISQILGQFA-----CMPPEHR 114
DB 26 CLIFHCVSDSRSVNRKTV-----KFNHTSVHGVGHSFVQSAFKAFXLVPEAVEQK 78
XX
QY 115 DYD 117
DB 79 DPD 81
XX
XX RESULT 12
XX ADB91448 standard; protein; 101 AA.
XX
XX ADB91448 standard; protein; 101 AA.
XX
XX ADB91448 standard; protein; 101 AA.
XX
XX 01-JAN-2004 (first entry)
XX
XX Human secreted protein - SEQ ID 977.
XX
XX antihaemic; antithrombotic; antiarthritic; antiinflammatory; antithyroid;
XX antidiabetic; immunosuppressive; dermatological; nephrotoxic;
XX antiparkinsonian; neuroprotective; neurotrophic; antibacterial; virocid;
XX fungicide; antiparasitic; antiarteriosclerotic; vulnary; cytostatic;
XX haemopoietic; haematologic; anaemia; autoimmune disorder; diabetes;
XX rheumatoid arthritis; inflammation; Grave's disease; diabetes;
XX systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
XX Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
XX cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
XX human.
XX
XX Homo sapiens.
XX
XX WO2003038063-A2.
XX
XX 08-MAY-2003.
XX
XX 19-MAR-2002; 2002WO-US008277.
```

```
PR 21-MAR-2001; 2001US-0277340P.
PR 19-JUL-2001; 2001US-0306171P.
PR 13-NOV-2001; 2001US-0331267P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX MPI; 2003-229407/22.
XX Nucleic acid encoding a human secreted protein is useful in diagnosing or
XX treating diabetes or conditions related to diabetes.
XX
PS Claim 3; SEQ ID NO 622; 1537pp; English.
XX
XX The invention relates to isolated nucleic acid molecules ADB91065-
XX ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-
XX ADB91834. Also disclosed is a recombinant vector comprising a
XX polynucleotide of the invention, and a recombinant host cell comprising
XX the recombinant vector. The polypeptide of the invention is useful in
XX identifying a binding partner by contacting the polypeptide with a
XX binding partner, and determining whether the binding partner increases or
XX decreases activity of the polypeptide. The polypeptide, polynucleotide,
XX antibody or its fragment, agonist or antagonist are useful for preparing
XX a pharmaceutical composition for diagnosing or treating diabetes or
XX conditions related to diabetes. The present sequence is that of the human
XX immunoglobulin Fc portion used to generate fusion proteins, increasing
XX the stability of the fused protein as compared to the secreted protein
XX only. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcc_sequences.
XX
SQ Sequence 101 AA;
XX
XX Query Match 7.7%; Score 71.5; DB 7; Length 101;
XX Best Local Similarity 27.0%; Pred. No. 3.3;
XX Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;
XX
QY 65 CLIIDIGNETELRLDPTSLGYEVQKFLHSMGISQILGQFA-----CMPPEHR 114
DB 26 CLIFHCVSDSRSVNRKTV-----KFNHTSVHGVGHSFVQSAFKAFXLVPEAVEQK 78
XX
QY 115 DYD 117
DB 79 DPD 81
XX
XX RESULT 13
XX ADB91448 standard; protein; 102 AA.
XX
XX ADB91448 standard; protein; 102 AA.
XX
XX ADB91448 standard; protein; 102 AA.
XX
XX 02-DEC-1999 (first entry)
XX
XX Human secreted protein encoded by gene 39 clone HTLSP53.
XX
XX antihaemic; antithrombotic; antiarthritic; antiinflammatory; antithyroid;
XX antidiabetic; immunosuppressive; dermatological; nephrotoxic;
XX antiparkinsonian; neuroprotective; neurotrophic; antibacterial; virocid;
XX fungicide; antiparasitic; antiarteriosclerotic; vulnary; cytostatic;
XX haemopoietic; haematologic; anaemia; autoimmune disorder; diabetes;
XX rheumatoid arthritis; inflammation; Grave's disease; diabetes;
XX systemic lupus erythematosus; glomerulonephritis; neurodegenerative;
XX Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;
XX cancer; bacterial; viral; fungal; parasitic infection; gene therapy;
XX human.
XX
XX Homo sapiens.
XX
XX WO947540-A1.
```

PD 23-SEP-1999.  
XX  
XX 18-MAR-1999; 99WO-US005804.  
XX  
XX 19-MAR-1998; 98US-0078563P.  
PR 19-MAR-1998; 98US-0078566P.  
PR 19-MAR-1998; 98US-0078573P.  
PR 19-MAR-1998; 98US-0078574P.  
PR 19-MAR-1998; 98US-0078576P.  
PR 19-MAR-1998; 98US-0078577P.  
PR 19-MAR-1998; 98US-0078578P.  
PR 19-MAR-1998; 98US-0078579P.  
PR 19-MAR-1998; 98US-0078581P.  
PR 01-APR-1998; 98US-0080312P.  
PR 01-APR-1998; 98US-0080313P.  
PR 01-APR-1998; 98US-0080314P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;  
PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
PI Shi Y, Moore PA;  
XX  
XX WPI; 1999-562050/47.  
DR N-PSDB; AA224849.  
XX  
XX New isolated human genes, useful for diagnosis and treatment of e.g.  
PT cancers, neurological disorders, immune diseases, inflammation or blood  
PT disorders.  
XX  
XX Claim 11; Page 382; 484p; English.  
XX  
XX This sequence represents a secreted human protein encoded by the gene  
CC clone detailed in the descriptor line. The gene can be used to generate  
CC fusion proteins by linking to the gene to a human immunoglobulin re-  
CC portion (e.g. AA224802) for increasing the stability of the fused protein  
CC as compared to the human protein only. The invention relates to 95 novel  
CC genes and their fragments (nucleic acid sequences: AA224811-224907; amino  
CC acid sequences AA41308-41404) which are useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
CC pathological conditions can be diagnosed by determining the amount of the  
CC new polypeptides in a sample or by determining the presence of mutations  
CC in the new polynucleotides. Specific uses are described for each of the  
CC 95 polynucleotides, based on which tissues they are most highly expressed  
CC in (see AA224811 for described uses)  
XX  
SQ Sequence 102 AA;  
Query Match 7.7%; Score 71.5; DB 2; Length 102;  
Best Local Similarity 27.0%; Pred. No. 3.4;  
Matches 17; Conservative 10; Mismatches 19; Indels 17; Gaps 2;  
QY 65 CLIIDCIGNETELRLDFTSLGYEYOKFLHLSMGISQILGQFA-----CMPEHR 114  
DQ 26 CIIFHCVSDRSRNVRETKV-----KPVHTSVHGVGSHSFVQSAKAPXLPPEAVBQK 78  
QY 115 DYD 117  
DQ 79 DPD 81  
Db 79 DPD 81  
RESULT 14  
ADB07674  
ID ADB07674 standard; protein; 133 AA.  
XX  
XX ADB07674;  
AC  
XX 20-NOV-2003 (first entry)  
DT  
XX  
DE Allolococcus otitidis antigenic protein SEQ ID NO:1614.  
XX  
XX Allolococcus otitidis; antigenic protein; immunogenic; immunisation;  
KM gene therapy; Gram-positive bacterium; infection.

XX  
XX Allolococcus otitidis.  
OS  
XX  
XX WO2003048304-A2.  
PN  
XX  
XX 12-JUN-2003.  
PD  
XX  
XX 25-NOV-2002; 2002WO-US036123.  
PF  
XX  
XX 29-NOV-2001; 2001US-0333777P.  
PR 18-NOV-2002; 2002US-0426742P.  
XX  
XX (AMHP) WYETH HOLDINGS CORP.  
PA  
XX  
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
PI  
PI WPI; 2003-505284/47.  
DR N-PSDB; ADB07673.  
XX  
XX  
XX Claim 33; SEQ ID NO 1614; 1019p; English.  
XX  
XX The present invention describes an isolated polynucleotide (1) of  
CC Allolococcus otitidis genomic DNA, which encodes an antigenic protein.  
CC Allolococcus otitidis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
CC expression vector comprising the novel isolated polynucleotide (1), its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against Allolococcus otitidis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying Allolococcus  
CC otitidis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Allolococcus  
CC otitidis. The present sequence represents an Allolococcus otitidis  
CC antigen protein from the present invention.  
XX  
SQ Sequence 133 AA;  
Query Match 7.3%; Score 67.5; DB 6; Length 133;  
Best Local Similarity 23.3%; Pred. No. 15;  
Matches 24; Conservative 24; Mismatches 36; Indels 19; Gaps 4;  
QY 9 TSYRVNLOAAIOLKIDBSNNFREBPYKSIQSEAFILPOSIPERYKMSKPLGICLII 68  
DQ 33 TNRPHAFQALLESORRPLKRY-----VRAFOSLERLYLTPIEASFHALNSGP-----I 82  
QY 69 DCIGNETELRLDFTSLGYEYOKFLH-----LSWHGISQILGQ 106  
DQ 83 EGMNNKTKTLKRT-----GYGRRPQHPRRYRILILNRLUTEKQ 121  
Db 83 EGMNNKTKTLKRT-----GYGRRPQHPRRYRILILNRLUTEKQ 121  
RESULT 15  
ABP31818  
ID ABP31818 standard; protein; 102 AA.  
XX  
XX ABP31818;  
AC  
XX  
XX



DT 08-JUL-2002 (first entry)  
DE Human ORF791 protein, SEQ ID NO:1582.  
XX  
XX Human; ORF, open reading frame; ORF, drug screening; diagnosis;  
KM disease monitoring; cytokine; cell proliferation; cell differentiation;  
KM immune modulation; haematopoiesis regulation; tissue growth;  
KM angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KM thrombolytic; tumour inhibition; bodily characteristics; fertility;  
KM behaviour; cancer; proliferative disorder; neurological disorder;  
KM cardiovascular disease; immune system disorder; organ transplantation;  
KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KM hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
KM vasotropic; antipsoriatic; antidiabetic; cytosolic; nocitopic;  
KM neuroprotective; antithrombotic; anticoagulant; thrombolytic;  
KM cardiant; hypotensive; antihypertoid; antiinflammatory; immunomodulator;  
KM dermatological; analgesic; virucide; antibacterial; fungicide.  
OS Homo sapiens.  
XX  
XX WO200190366-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US017076.  
XX  
XX 24-MAY-2000; 2000US-0206690P.  
XX  
XX (CUPRA-) CURAGEN CORP.  
XX  
XX Leach MD, Shinkens RA;  
XX  
XX WPI, 2002-106200/14.  
XX  
XX N-PSDB; ABN75844.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation.  
XX  
XX Claim 10, Page 661; 2508BP; English.  
XX  
XX Sequences ABP1028-ABP35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
CC ABN75887 represent cDNAs encoding them. The invention also encompasses  
CC polypeptides at least 80% identical to the ORF-ORF4534 (collectively  
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or  
CC activity, and methods of screening individuals for a predisposition to an  
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
CC range of biological activities, such as cytokine, cell proliferation,  
CC cell differentiation, immune modulation, haematopoiesis regulation,  
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
CC and antiinfective activity, and may also be involved in the determination  
CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
CC nucleic acids and antibodies may be used in the treatment of cancers,  
CC other proliferative disorders such as psoriasis and benign tumours,  
CC neurological disorders such as epilepsy and Alzheimer's disease,  
CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes, in the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX

CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases  
XX

Seq Sequence 102 AA;

Query Match 7.2%; Score 66.5; DB 5; Length 102;

Best Local Similarity 21.4%; Pred. No. 14;

Matches 21; Conservative 21; Mismatches 33; Indels 23; Gaps 4;

QY 2 OSVCGAGTSTYNNVLOALIQSLKDPNNPFEPPYKSIQSEAFLLPQSTPEERYKMKSK- 60

Db 16 QPISNGITSHNL-----NNPGR--DSRAEKVMIFRRDIPKXKQROCTSE 60

QY 61 -PLGICLIID-----CIGNETELLRDPTSLGYEVQK 91

Db 61 HPFGTIRKHYDAGYFLCKGNEKVTAEYVALSCLGYDIR 98

Search completed: September 9, 2004, 06:09:30

Job time : 59 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: September 9, 2004, 06:09:34 : Search time 54 Seconds

(without alignments)  
1051.149 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252

Perfect score: 930

Sequence: 1 KQSVGAGTXYRNVLAIAIQ.....YLAKPKMFIQNVVSDGQ 177

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 737077

Minimum DB seq length: 0

Maximum DB seq length: 177

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/FCI\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCMTS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	166	17.8	147	14	US-10-214-932-110
2	154	16.6	167	9	US-09-864-761-48728
3	130	14.0	163	9	US-09-864-761-47950
4	108	11.6	158	10	US-09-851-873-61
5	78	8.4	134	15	US-10-114-432-32
6	78	8.4	134	15	US-10-114-432-33
7	78	8.4	146	15	US-10-114-432-7
8	78	8.4	146	15	US-10-114-432-20
9	78	8.4	146	15	US-10-114-432-65
10	78	8.4	146	15	US-10-114-432-70
11	78	8.4	174	15	US-10-114-432-34
12	77	8.3	52	9	US-09-989-903-51
13	77	8.3	52	14	US-10-068-564-51
14	76	8.2	56	9	US-09-989-903-58
15	76	8.2	56	14	US-10-068-564-58

16	75.5	8.1	74	9	US-09-989-903-18	Sequence 18, App1
17	75.5	8.1	74	14	US-10-068-564-18	Sequence 18, App1
18	75.5	8.1	77	9	US-09-989-903-25	Sequence 25, App1
19	75.5	8.1	77	14	US-10-068-564-25	Sequence 25, App1
20	74.5	8.0	94	9	US-09-989-903-14	Sequence 14, App1
21	74.5	8.0	94	14	US-10-068-564-14	Sequence 14, App1
22	73	7.8	39	8	US-08-459-455-45	Sequence 45, App1
23	71.5	7.7	101	12	US-10-653-595-149	Sequence 149, App1
24	71.5	7.7	102	12	US-09-397-945-149	Sequence 149, App1
25	70	7.5	51	9	US-09-989-903-45	Sequence 45, App1
26	70	7.5	51	14	US-10-068-564-45	Sequence 45, App1
27	68	7.3	39	9	US-09-989-903-38	Sequence 38, App1
28	68	7.3	39	14	US-10-068-564-38	Sequence 38, App1
29	66.5	7.2	102	11	US-09-864-408A-1582	Sequence 1582, App1
30	66	7.1	39	9	US-09-989-903-30	Sequence 30, App1
31	66	7.1	39	14	US-10-068-564-30	Sequence 30, App1
32	66	7.1	167	15	US-10-429-872-20	Sequence 20, App1
33	65.5	7.0	130	12	US-10-424-599-271119	Sequence 271119, App1
34	64.5	6.9	141	12	US-10-424-599-231330	Sequence 231330, App1
35	64.5	6.9	174	12	US-10-282-122A-76633	Sequence 76633, App1
36	64	6.9	28	9	US-09-989-903-50	Sequence 50, App1
37	64	6.9	28	14	US-10-068-564-50	Sequence 50, App1
38	63.5	6.8	166	12	US-10-282-122A-73139	Sequence 73139, App1
39	63.5	6.8	171	12	US-10-282-122A-76190	Sequence 76190, App1
40	63	6.8	128	12	US-10-424-599-178064	Sequence 178064, App1
41	63	6.8	153	9	US-09-942-446-7	Sequence 7, App1
42	62.5	6.7	110	16	US-10-437-963-120464	Sequence 120464, App1
43	62.5	6.7	128	16	US-10-437-963-193788	Sequence 193788, App1
44	62.5	6.7	167	14	US-10-128-714-8488	Sequence 8488, App1
45	62	6.7	140	15	US-10-289-762-888	Sequence 888, App1

#### ALIGNMENTS

RESULT 1  
US-10-214-932-110  
; Sequence 110, Application US/10214932  
; Publication No. US20030100707A1  
; GENERAL INFORMATION:  
; APPLICANT: HWANG, Inhan  
; APPLICANT: KIM, Dae Heon  
; APPLICANT: LEE, Yong Jik  
; TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE  
; FILE REFERENCE: APOB2/US  
; CURRENT APPLICATION NUMBER: US/10/214,932  
; CURRENT FILING DATE: 2002-08-08  
; NUMBER OF SEQ ID NOS: 133  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 110  
; LENGTH: 147  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-214-932-110

Query Match 17.8%; Score 166; DB 14; Length 147;  
Best Local Similarity 32.6%; Pred. No. 2.2e-10;  
Matches 43; Conservative 22; Mismatches 49; Indels 18; Gaps 4;

QY	52	EERYKXKXKXGICIID-----CIGNETEL--LRDTFTSICGVQGFHLISM	97
DB	6	DNSYKMDYEMWGLIINNKNFKSTGWTSGTDVDANLRFTFRNLKYEVRKNLDLNR	65
QY	98	HGISQLGQFACMPHRRYDSFVCLVSRGSGQSVYGVDTHTSLPLPHHRRMMPDSCP	157
DB	66	BEIVELMRDVS-KEDHSKRSSTFVCLLSHGEGGIIFG---TNGFVDLKKLTINFRGRCR	121
QY	158	YLAKPKMFFIQ 169	
DB	122	SLTGKPKLFIQ 133	

RESULT 2

US-09-864-761-48728  
; Sequence 48728, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 48728  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007256.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
; OTHER INFORMATION: EST HUMAN HIT: BE271526.1, EVALUE 2.00e-98  
; OTHER INFORMATION: SWISSPROT HIT: Q14790, EVALUE 2.00e-99  
US-09-864-761-48728

Query Match 16.6%; Score 154; DB 9; Length 167;  
Best Local Similarity 35.4%; Pred. No. 6.1e-09;  
Matches 35; Conservative 18; Mismatches 38; Indels 8; Gaps 3;

QY 78 LRDTFTSLGYEVQKFLHLSHGISQILGFCACPEHRDVSFVCVLSRGSQSVYGVQD 137  
DB 2 LTTTFELHFEIKPHDDCTVEQIYELIKIYOLM-DHSNMDCIFCCLSHGKGIYGDG 60  
QY 138 THSGPLPHHTRKRFMDSCPYLAGKPKMFFIQ-----NY 171  
DB 61 QEA--PIYELTSGTGLKCPISLAGKPKVFIOACGSDNY 97

RESULT 3  
US-09-864-761-47950  
; Sequence 47950, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Hank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 47950  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007283.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
; OTHER INFORMATION: EST HUMAN HIT: BE271526.1, EVALUE 4.00e-94  
; OTHER INFORMATION: SWISSPROT HIT: Q92851, EVALUE 3.00e-94  
US-09-864-761-47950

Query Match 14.0%; Score 130; DB 9; Length 163;  
Best Local Similarity 33.3%; Pred. No. 3e-06;  
Matches 31; Conservative 17; Mismatches 43; Indels 2; Gaps 1;

QY 77 LRDTFTSLGYEVQKFLHLSHGISQILGFCACPEHRDVSFVCVLSRGSQSVYGVQD 136  
DB 1 ILTSHVQWLGFVTHINNVTVKEMEMVLQOKCNPAHADDCFFVFCILTHGRGAVYSSD 60

```

QY      137 QTHSGLELHIRRMFMWDSCPYLAGKRPKMFIIQ 169
      : : : | | | | | : | | |
Db      61 E--ALIPIREIMSHPTALQCPRLAEKPLFFIIQ 91

```

RESULT 4  
US-09-85

```

1 Sequence 61, Application US/09851-873
2 Publication No. US20030165488A1
3
4 GENERAL INFORMATION:
5 APPLICANT: Kietzlen, Rolf F
6 APPLICANT: Reacton, Ilene M
7 APPLICANT: Welland, Katherine L
8 TITLE OF INVENTION: HUMAN CASPASE-12 MATE
9 FILE REFERENCE: 283411/0023
10 CURRENT APPLICATION NUMBER: US/09/851,873
11 CURRENT FILING DATE: 2001-05-08
12 NUMBER OF SEQ ID NOS: 105
13
14 SOFTWARE: PatentIn Ver. 2.0
15
16 SEQ ID NO: 61
17 LENGTH: 158
18 TYPE: PR
19 ORGANISM: Homo sapiens
20
21 US-09-851-873-61

```

Query Match	11.6%;	Score 108;	DB 10;	Length 158;
Best Local Similarity	26.4%;	Pred. No. 0.00089;		
Matches 42;	Conservative 19;	Mismatches 70;	Indels 28;	Gaps 4;

```
QY      26 PSNNRREPVKSIQSEAFLPQSIPERRKMKSLGLTIIDC-----IGNE 74
```

DQ 75 TELL-RDTFTSLGVEYQKFLHLSMHGSQLQGACMPHRHYDSFCVVLVSNGSGSQSVY 133  
 : ||| : | : ||| ||| : ||| :  
 Db 58 LDIIGMDLLENLGYSVVIKESITAGEMETALRQFAAPHEQSDDSTFVFWMSHILNIGIC 117

QY 134 GV--DQTHSGPLNHRMFEWGSCEYLAKPKMFIQ 169  
| | | | : | | | : |  
Db 118 GTGHWDEPDVLHDDTIFEIPNNRNCQSLKDKPKVILMQ 156

RESULT 5  
10 1 1 4 33 33

```

US-10-114-432-32
Sequence 32, Application US/10114432
Publication No. US20040015915A1
GENERAL INFORMATION:
APPLICANT: Chaillita-Bid, Pia M.
APPLICANT: Raitano, Arthur B.
APPLICANT: Paris, Mary
APPLICANT: Hubbert, Rene S.
APPLICANT: Morrison, Robert K.
APPLICANT: Ge, Mangmao
APPLICANT: Jakobovits, Aya
TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
TITLE OF INVENTION: ENTITLED 213P.F.1 USEFUL IN TREATMENT AND
TITLE OF INVENTION: DETECTED OF CANCER
FILE REFERENCE: 51158-20066.00
CURRENT APPLICATION NUMBER: US/10/114.432
CURRENT FILING DATE: 2002-04-01
NUMBER OF SEQ ID NOS: 88
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 32
LENGTH: 134
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-114-432-32

```

Query Match 8.4%; Score 78; DB 15; Length 134;  
Best Local Similarity 25.6%; Pred. No. 1.8;  
Matches 34; Conservative 22; Mismatches 61; Indels 16; Gaps 6;

```
QY 47 PGIPIPERYKMSKPLGICILIIDCI-----GNRETL--LROTPISIGYEQKPHLSMHG 99
Db 4 PRSIEEEKQDMSCARLALIL-----CVTKAREGSEEDLALHHPFQRLPRESTMKRDPTEQ 60
QY 100 ISQILGQFACMPEHRDYDSFVC---VLSRGSQSVYGVQDTHSGLPHTHIRMFMGDSG 156
Db 61 FQEBLEKFGQALDSR-DEVSCAFVYLAHGREGLKGED--GENVKLENLEALNNNGC 11
QY 157 PYLAGKPKVEFIQ 169
Db 118 QALRAKPKYVILQ 130
```

RESULT  
US-10-1

```

? Sequence 33, Application US/20114432
? Publication No. US20040019915A1
? GENERAL INFORMATION:
? APPLICANT: Challinor-Bid, Pia M.
? APPLICANT: Raitano, Arthur B.
? APPLICANT: Paris, Mary
? APPLICANT: Hubert, Rene S.
? APPLICANT: Morrison, Robert K.
? APPLICANT: Ge, Wangmao
? APPLICANT: Jakobivics, Aya
? TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
? TITLE OF INVENTION: ENTITLED 213P11 USEFUL IN TREATMENT AND
? TITLE OF INVENTION: DETECTION OF CANCER
? FILE REFERENCE: 51158-20066.00
? CURRENT APPLICATION NUMBER: US/10/114,432
? CURRENT FILING DATE: 2002-04-01
? NUMBER OF SEQ ID NOS: 88
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 33
? LENGTH: 134
? TYPE: PRT
? ORGANISM: Homo Sapiens
US-10-114-432-53

```

Query Match	8.4%;	Score 78;	DB 15;	Length 134;
Best Local Similarity	25.6%;	Pred. No. 1.8;		
Matches	34	W:matches	61	Indels 16
				Gap 5

```

QY 47 POSTPERKYMKSKPLGICLLIDCI-----GNETEL--LRDTFSLGYEVQKFLHLSMHG 99
DB 4 PRSIEEKYKDMSCARLALLI---CVTKAREGBEBDLALEHMFOLPSTETMKRDPABEQ 60
QY 100 ISOLIGFACMPHEHDYDSFVC---VLVSRGSGSVYGVDPDTHSGPLPHIRRMFMDSC 156
DB 61 FOEBLEFFQOALISRS-DEVSCAFVLLAHGREGLKGED--GEMXKLENIFFALNNKNC 117
QY 157 PYLAGKPMFFIQ 169
DB 118 QALRAKRPVYIIQ 130

```

## RESULT 7

US-10-114,432-7  
 ? Sequence 7, Application US/10114432  
 ? Publication No. US20040019915A1  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Challita-Eid, Pia M.  
 ? APPLICANT: Raitano, Arthur B.  
 ? APPLICANT: Paris, Mary  
 ? APPLICANT: Hubert, Rene S.  
 ? APPLICANT: Morrison, Robert K.  
 ? APPLICANT: Ge, Wangmao  
 ? APPLICANT: Jakobovits, Aya  
 ? TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN  
 ? TITLE OF INVENTION: ENTITLED 213P1I USEFUL IN TREATMENT AND  
 ? TITLE OF INVENTION: DETECTION OF CANCER  
 ? FILE REFERENCE: 51158-20066.00  
 ? CURRENT APPLICATION NUMBER: US/10/114,432

```

; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Homo Sapiens
US-10-114-432-7
```

```

Query Match      8.4%; Score 78; DB 15; Length 146;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 34; Conservative 22; Mismatches 61; Indels 16; Gaps 6;
```

```

QY 47 POSIPBERRYKMSKPLGICLIIDCI-----GNTEL--LRDTFTSLGYEVQKFLHLSMHG 99
DB 4 PRSLBEEKIDMSGARLAIL--CVTKAREGSEEDLDALHMFROLFSTMKRDPYAEQ 60
QY 100 ISOILGQFACMPBHRDYDSFVC---VLVSRGSGSVYGVQDTHSGPLPHHIRMFMGDSG 156
DB 61 FOEELKFFQQAIDSRE-DPVSCAFVVLMAHGRGFLKGED--GEWVKLENLFEALNNKNC 117
QY 157 PYLAGKPKMFFIQ 169
DB 118 QALRAKPKVYIIQ 130
```

```

RESULT 8
US-10-114-432-20
; Sequence 20, Application US/10114432
; Publication No. US20040019915A1
; GENERAL INFORMATION:
; APPLICANT: Chailita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Mangmao
; APPLICANT: Jakobivits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; TITLE OF INVENTION: ENTITLED 213P1F11 USEFUL IN TREATMENT AND
; FILE REFERENCE: 51158-20066.00
; CURRENT APPLICATION NUMBER: US/10/114,432
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Homo Sapiens
US-10-114-432-20
```

```

Query Match      8.4%; Score 78; DB 15; Length 146;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 34; Conservative 22; Mismatches 61; Indels 16; Gaps 6;
```

```

QY 47 POSIPBERRYKMSKPLGICLIIDCI-----GNTEL--LRDTFTSLGYEVQKFLHLSMHG 99
DB 4 PRSLBEEKIDMSGARLAIL--CVTKAREGSEEDLDALHMFROLFSTMKRDPYAEQ 60
QY 100 ISOILGQFACMPBHRDYDSFVC---VLVSRGSGSVYGVQDTHSGPLPHHIRMFMGDSG 156
DB 61 FOEELKFFQQAIDSRE-DPVSCAFVVLMAHGRGFLKGED--GEWVKLENLFEALNNKNC 117
QY 157 PYLAGKPKMFFIQ 169
DB 118 QALRAKPKVYIIQ 130
```

```

RESULT 9
US-10-114-432-65
; Sequence 65, Application US/10114432
; Publication No. US20040019915A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Chailita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Mangmao
; APPLICANT: Jakobivits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; TITLE OF INVENTION: ENTITLED 213P1F11 USEFUL IN TREATMENT AND
; FILE REFERENCE: 51158-20066.00
; CURRENT APPLICATION NUMBER: US/10/114,432
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Homo Sapiens
US-10-114-432-65
```

```

Query Match      8.4%; Score 78; DB 15; Length 146;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 34; Conservative 22; Mismatches 61; Indels 16; Gaps 6;
```

```

QY 47 POSIPBERRYKMSKPLGICLIIDCI-----GNTEL--LRDTFTSLGYEVQKFLHLSMHG 99
DB 4 PRSLBEEKIDMSGARLAIL--CVTKAREGSEEDLDALHMFROLFSTMKRDPYAEQ 60
QY 100 ISOILGQFACMPBHRDYDSFVC---VLVSRGSGSVYGVQDTHSGPLPHHIRMFMGDSG 156
DB 61 FOEELKFFQQAIDSRE-DPVSCAFVVLMAHGRGFLKGED--GEWVKLENLFEALNNKNC 117
QY 157 PYLAGKPKMFFIQ 169
DB 118 QALRAKPKVYIIQ 130
```

```

RESULT 10
US-10-114-432-70
; Sequence 70, Application US/10114432
; Publication No. US20040019915A1
; GENERAL INFORMATION:
; APPLICANT: Chailita-Eid, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Robert K.
; APPLICANT: Ge, Mangmao
; APPLICANT: Jakobivits, Aya
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
; TITLE OF INVENTION: ENTITLED 213P1F11 USEFUL IN TREATMENT AND
; FILE REFERENCE: 51158-20066.00
; CURRENT APPLICATION NUMBER: US/10/114,432
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Homo Sapiens
US-10-114-432-70
```

```

Query Match      8.4%; Score 78; DB 15; Length 146;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 34; Conservative 22; Mismatches 61; Indels 16; Gaps 6;
```

```

QY 47 POSIPBERRYKMSKPLGICLIIDCI-----GNTEL--LRDTFTSLGYEVQKFLHLSMHG 99
DB 4 PRSLBEEKIDMSGARLAIL--CVTKAREGSEEDLDALHMFROLFSTMKRDPYAEQ 60
```

QY 100 ISQILGQFCAMPEHNDHYSFVC---VLSRGSQS VVGVDTHSGLPYHHRKAFMDS C 156  
 Db 61 FQSELEKFGQALDSE-DPVSCAFVYLAHAREGLKGD--GMVYKLENIFFALNNKC 117  
 QY 157 PYLAGKPKMFTIQ 169  
 Db 118 QALRAKPKVITIQ 130

```

RESULT 11
US-10-114-432-34
; Sequence 34. Application US/10114432
; Publication No. US20040019915A1
GENERAL INFORMATION:
APPLICANT: Chailita-Eid, Pia M.
APPLICANT: Raitano, Arthur B.
APPLICANT: Paris, Mary
APPLICANT: Hubert, Rene S.
APPLICANT: Morrison, Robert K.
APPLICANT: Ge, Wangmao
APPLICANT: Jakobivits, Aya
TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN
TITLE OF INVENTION: ENTITLED 213p1.1 USEFUL IN TREATMENT AND
TITLE OF INVENTION: DETECTION OF CANCER
FILE REFERENCE: 51158-20065.00
CURRENT APPLICATION NUMBER: US/10/114,432
CURRENT FILING DATE: 2002-04-01
NUMBER OF SEQ ID NOS: 88
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 174
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-114-432-34

```

Query Match	8.4%;	Score 78;	DB 15;	Length 174;
Best Local Similarity	25.6%;	Pred. No. 2.5;		
Matches	34;	Conservative	22;	Mismatches 61;
			Indels	16;
			Gaps	6;
QY	47	POSIPERRYKMSKPLGICLIIDCT---- <td>59</td> <td></td>	59	
	:	:	:	
Db	4	PRLEEEKXDMGSGARLATL--CVTKARSESEEDLALHEMRQRLFEETWMRDPTEAQ	60	
	:	:	:	
QY	100	ISGLIQCFPCMEHRDYDSFVC--VLVSRGSGSYGVQDTHSGPLHHIRKMPMGDSC	156	
	:	:	:	
Db	61	FOEELERKFOAADSRB-DEVSCAFVLMAGHREGFLKGED--GEKVLTELFEALNNKNC	117	
	:	:	:	
QY	157	PVLAGEKPMFFIQ	169	
	:	:	:	
Db	118	QALRAKPKVYIIQ	130	

```

RESULT 12
US-09-989-903-51
: Sequence 51, Application US/09989903
: Patent No. US20020146804A1
: GENERAL INFORMATION:
: APPLICANT: Alnemri, Emdad S.
: TITLE OF INVENTION: Fernandez-Alnemri, Teresa
: TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
: TITLE OF INVENTION: AND METHODS OF USE
: PRT REFERENCE: 460140, 434D1
: CURRENT APPLICATION NUMBER: US/09/989,903
: CURRENT FILING DATE: 2002-04-11
: NUMBER OF SEQ ID NOS: 78
: SOFTWARE: PatentIn ver. 2.0
: SEQ ID NO 51
: LENGTH: 52
: TYPE: PRT
: ORGANISM: Mus musculus
US-09-989-903-51

```

Query Match 8.3%; Score 77; DB 9; Length 52;

```

QY      144 LHHRRMFWDSCPYLACKPRMFFIQ 169
      :::|::|::|::|::|::|::|::|
Db      4 IYDLISYFGSKCPRLSCKPKRFFIQ 29

```

```

; RESULT 13
; US-10-068-564-51
; Sequence 51, Application US/10068564
; Publication No. US20030040096A1
;
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: 480140.434C2
; CURRENT APPLICATION NUMBER: US/10/068,564
; CURRENT FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 78
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 51
; LENGTH: 52
; TYPE: PR1
; ORGANISM: Mus musculus
;
; US-10-068-564-51

```

```

Query Match      8.3%; Score 77; DB 14; Length 52;
Best Local Similarity 50.0%; Pred. No. 0.6;
Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY      144 LHHIRRMFGDSCPTLACKPKPFITQ 169
      ::::|::|::|::|::|::|::|::|
Db      4 IYDLTNYFTGSKCPKPLSGKPKFIITQ 29

```

```

US-09-389-903-58
RESULT 14
; Sequence 58, Application US/09989903
; Patent No. US20020146804v1
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emed S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-11, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: 480140.434D1
; CURRENT APPLICATION NUMBER: US/09/989,903
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 56
; TYPE: FRT
; ORGANISM: Mus musculus
US-09-969-903-58

```

Query Match	8.2%;	Score 76;	DB 9;	Length 56;
Best Local Similarity	53.6%;	Pred. No. 0.87;		
Matches	15;	Conservative	3;	Mismatches 10;
				Indels 0;
				Gaps 0;

Cy 142 LPLHRRMFMGDSCPYLAGKPKMFIIQ 169  
| : | : |||||  
Db 2 LQLQEVFRLFDNANCPSLQNKPKMFIIQ 29

RESULT 15  
US-10-068-564-58  
; Sequence 58, Application US/10068564  
; Publication No. US20030040096A1  
; GENERAL INFORMATION:  
; APPLICANT: Alnemri, Emed S.  
; APPLICANT: Fernandez-Alnemri, Teresa

```

; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: 480140.434C2
; CURRENT APPLICATION NUMBER: US/10/068,564
; CURRENT FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 58
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-068-564-58

```

```

Query Match      8.2%  Score 76; DB 14; Length 56;
Best Local Similarity 53.6%; Pred. No. 0.87;
Matches 15; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

```

```

Oy      142 LPLHRRMFMGDCPYLAGKPKXFFIQ 169
Db      2 LQLQEVFRLFDNANCPSLQXKPKXFFIQ 29

```

Search completed: September 9, 2004, 06:12:22  
 Job time : 61 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:08:40 ; Search time 18 Seconds

(without alignments)  
507.655 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_76\_252

Sequence: 1 KOSVCGAGTSYRNVLQALIC.....YLAKRPMFPIQNVVSDGQ 177

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 239862

Minimum DB seq length: 0  
Maximum DB seq length: 177

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cg2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cg2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cg2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cg2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cg2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cg2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	166	17.8	148	3	US-08-964-308-11
2	166	17.8	148	3	US-08-964-313-11
3	166	17.8	148	4	US-09-069-138-11
4	77	8.3	52	4	US-09-187-789-51
5	77	8.3	52	4	US-09-139-600-46
6	76	8.2	56	4	US-09-187-789-58
7	76	8.2	56	4	US-09-139-600-53
8	75.5	8.1	74	4	US-09-187-789-18
9	75.5	8.1	74	4	US-09-139-600-13
10	75.5	8.1	77	4	US-09-187-789-25
11	75.5	8.1	77	4	US-09-139-600-20
12	74.5	8.0	94	4	US-09-187-789-14
13	74.5	8.0	94	4	US-09-139-600-9
14	73	7.8	39	3	US-08-258-287B-46
15	73	7.8	39	3	US-08-368-704C-45
16	70.5	7.6	173	4	US-09-540-236-3337
17	70.5	7.5	51	4	US-09-187-789-45
18	70	7.5	51	4	US-09-139-600-40
19	68	7.3	39	4	US-09-187-789-38
20	68	7.3	39	4	US-09-139-600-33
21	66	7.1	39	4	US-09-187-789-30
22	66	7.1	39	4	US-09-139-600-25
23	64	6.9	28	4	US-09-187-789-50
24	64	6.9	28	4	US-09-139-600-45
25	62	6.7	140	4	US-09-158-452A-888
26	62	6.7	161	4	US-09-328-352-4309
27	60	6.5	28	4	US-09-187-789-44

28	60	6.5	28	4	US-09-139-600-39	Sequence 39, Appl
29	59.5	6.4	161	4	US-09-252-991A-32451	Sequence 32451, A
30	58.5	6.3	126	4	US-09-134-001C-3766	Sequence 3766, Ap
31	58	6.2	28	4	US-09-187-789-29	Sequence 29, Appl
32	58	6.2	28	4	US-09-139-600-24	Sequence 24, Appl
33	58	6.2	97	1	US-08-204-740-7	Sequence 7, Appl
34	58	6.2	97	3	US-09-081-167A-7	Sequence 7, Appl
35	58	6.2	97	3	US-09-081-395-7	Sequence 7, Appl
36	58	6.2	97	3	US-09-416-833-7	Sequence 7, Appl
37	58	6.2	97	5	PCT-US95-02521-7	Sequence 7, Appl
38	58	6.2	135	4	US-09-134-001C-4301	Sequence 4301, Ap
39	57	6.1	106	4	US-09-328-352-7954	Sequence 7954, Ap
40	57	6.1	154	4	US-09-387-418A-11	Sequence 11, Appl
41	57	6.1	157	4	US-09-252-991A-25900	Sequence 25900, A
42	56.5	6.1	140	4	US-09-621-976-7010	Sequence 7010, Ap
43	56	6.0	95	4	US-09-732-210-740	Sequence 740, App
44	56	6.0	171	2	US-08-609-049A-20	Sequence 20, Appl
45	56	6.0	171	3	US-09-170-996-20	Sequence 20, Appl

## ALIGNMENTS

```
RESULT 1
US-08-964-308-11
; Sequence 11, Application US/08964308
; Patent No. 6066715
; GENERAL INFORMATION:
; APPLICANT: DESMARAIS, SYLVIE
; APPLICANT: FRIESEN, RICHARD
; APPLICANT: ZAMBONI, ROBERT
; TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,308
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: NORTH, ROBERT J
; REGISTRATION NUMBER: 27,366
; REFERENCE/DOCKET NUMBER: 19840 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-7262
; TELEFAX: 732-594-4720
; TEXT:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 148 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-964-308-11

Query Match 17.8%; Score 166; DB 3; Length 148;
Best Local Similarity 32.6%; Pred. No. 1.5e-12;
Matches 43; Conservative 22; Mismatches 49; Indels 18; Gaps 4;
QY 52 ERRYKMSKPLGICILID-----CIGNMETL--LRDTFTSLGYGVQKFLHSM 97
```

Db 7 DNSYKMDYPEMGKLCIIINNKCFHKSCTGMSRSGTDVDAANLRETRNLKYEVANRNDLTR 66  
Qy 98 HGISQILGQFACPKPEHRDYSFVCVLVSRGSGSVYGVDTGSHGLPLHHRMFMGDSGP 157  
Db 67 EEIYELMRDVS-KEDHSKRSSFVCVLLSHGEEGIIIFG--TNGPVDLKKITNFFRGDRCR 122  
Qy 158 YLAGKPKMFFIQ 169  
Db 123 SLTGKPKLFIIO 134

RESULT 2  
US-08-964-313-11  
Sequence 11, Application US/08964313  
Patent No. 6114132  
GENERAL INFORMATION:  
APPLICANT: DESMARAIS, SYLVIE  
APPLICANT: FRIESEN, RICHARD  
APPLICANT: GERESER, MICHAEL  
APPLICANT: KENNEDY, BRIAN  
APPLICANT: NICHOLSON, DONALD  
APPLICANT: RAMACHANDRAN, CHIDAMBARAN  
APPLICANT: SKOREY, KATHRYN  
APPLICANT: FORD-HUTCHINSON, ANTHONY  
TITLE OF INVENTION: PHOSPHATASE BINDING ASSAY  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PHILIPPE L. DURETTE - MERCK & CO., INC.  
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,313  
FILING DATE: 04-NOV-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 66/030,408  
FILING DATE: 04-NOV-1996  
APPLICATION NUMBER: PCT/CA97/00825  
FILING DATE: 03-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: DURETTE, PHILIPPE L.  
REGISTRATION NUMBER: 35,125  
REFERENCE/DOCKET NUMBER: 19824Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-4568  
TELEFAX: 732-594-4720  
TELEX:  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-964-313-11

Query Match 17.8%; Score 166; DB 3; Length 148;  
Best Local Similarity 32.6%; Pred. No. 1,5e-12;  
Matches 43; Conservative 22; Mismatches 49; Indels 18; Gaps 4;

Qy 52 EERYKMKSKPLGICLIID-----CIGNETEL--LRDTFTSLGYEVOKFLHLSM 97  
Db 7 DNSYKMDYPEMGKLCIIINNKCFHKSCTGMSRSGTDVDAANLRETRNLKYEVANRNDLTR 66  
Qy 98 HGISQILGQFACPKPEHRDYSFVCVLVSRGSGSVYGVDTGSHGLPLHHRMFMGDSGP 157  
Db 123 SLTGKPKLFIIO 134

Db 67 EEIYELMRDVS-KEDHSKRSSFVCVLLSHGEEGIIIFG--TNGPVDLKKITNFFRGDRCR 122  
Qy 158 YLAGKPKMFFIQ 169  
Db 123 SLTGKPKLFIIO 134

RESULT 3  
US-09-069-138-11  
Sequence 11, Application US/09069138  
Patent No. 6348572  
GENERAL INFORMATION:  
APPLICANT: DESMARAIS, SYLVIE  
APPLICANT: DURETTE, CLAUDE  
APPLICANT: FRIESEN, RICHARD  
APPLICANT: LEBLANC, YVES  
APPLICANT: ROY, PATRICK  
APPLICANT: YOUNG, ROBERT N.  
APPLICANT: ZAVBONI, ROBERT  
TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PHILIPPE L. DURETTE - MERCK & CO., INC.  
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/069,138  
FILING DATE: 29-APR-1998  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: DURETTE, PHILIPPE L.  
REGISTRATION NUMBER: 35,125  
REFERENCE/DOCKET NUMBER: 19840YIA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-4568  
TELEFAX: 732-594-4720  
TELEX:  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 148 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-069-138-11

Query Match 17.8%; Score 166; DB 4; Length 148;  
Best Local Similarity 32.6%; Pred. No. 1,5e-12;  
Matches 43; Conservative 22; Mismatches 49; Indels 18; Gaps 4;

Qy 52 EERYKMKSKPLGICLIID-----CIGNETEL--LRDTFTSLGYEVOKFLHLSM 97  
Db 7 DNSYKMDYPEMGKLCIIINNKCFHKSCTGMSRSGTDVDAANLRETRNLKYEVANRNDLTR 66  
Qy 98 HGISQILGQFACPKPEHRDYSFVCVLVSRGSGSVYGVDTGSHGLPLHHRMFMGDSGP 157  
Db 67 EEIYELMRDVS-KEDHSKRSSFVCVLLSHGEEGIIIFG--TNGPVDLKKITNFFRGDRCR 122  
Qy 158 YLAGKPKMFFIQ 169  
Db 123 SLTGKPKLFIIO 134

```
RESULT 4
US-09-187-789-51
; Sequence 51, Application US/09187789
; Patent No. 6340740
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434C1
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-187-789-51

Query Match
Best Local Similarity 8.3%; Score 77; DB 4; Length 52;
Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 144 LHHIRRMFGDSCPYLAGKPKMFFIQ 169
Db 4 IYDLTSYFTGSKCPSLSGKPKIFFIQ 29

RESULT 5
US-09-139-600-46
; Sequence 46, Application US/09139600
; Patent No. 6432628
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434
; CURRENT FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-139-600-46

Query Match
Best Local Similarity 8.3%; Score 77; DB 4; Length 52;
Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 144 LHHIRRMFGDSCPYLAGKPKMFFIQ 169
Db 4 IYDLTSYFTGSKCPSLSGKPKIFFIQ 29

RESULT 6
US-09-187-789-58
; Sequence 58, Application US/09187789
; Patent No. 6340740
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434C1
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
```

```
LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-187-789-58

Query Match
Best Local Similarity 8.2%; Score 76; DB 4; Length 56;
Matches 15; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 142 LPHHIRMFGDSCPYLAGKPKMFFIQ 169
Db 2 LQLOEVFRFLFDNANCPSLONKPKMFFIQ 29

RESULT 7
US-09-139-600-53
; Sequence 53, Application US/09139600
; Patent No. 6432628
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434
; CURRENT FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-139-600-53

Query Match
Best Local Similarity 8.2%; Score 76; DB 4; Length 56;
Matches 15; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 142 LPHHIRMFGDSCPYLAGKPKMFFIQ 169
Db 2 LQLOEVFRFLFDNANCPSLONKPKMFFIQ 29

RESULT 8
US-09-187-789-18
; Sequence 18, Application US/09187789
; Patent No. 6340740
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434C1
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-187-789-18

Query Match
Best Local Similarity 8.1%; Score 75.5; DB 4; Length 74;
Matches 19; Conservative 11; Mismatches 29; Indels 3; Gaps 1;

QY 111 PEHRDYDSFVCLVLRSGSGSVGVDTGSLPLH---TRRMFGDSCPYLAGKPKMFF 167
Db 1 PEHGSDSFTLVFWSHGLBEGICVKKRNKPKPVYLDHDTIFKIFNNSNCRSLNKKRILL 60

QY 168 IQ 169
```





This page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:11:06; Search time 15 seconds

(without alignments)  
615.626 Million cell updates/sec

Title: US-10-713-208-6\_Copy\_253\_348

Perfect score: 497

Sequence: 1 LEDSLLEVDGPMKKNVEFK.....ITSEKDHFFSLGCLLDVL 96

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 34992

Minimum DB seq length: 0

Maximum DB seq length: 96

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

PIR\_70: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	10.1	34	2 S44828	F54F2.3 protein -
2	49	9.9	74	2 AG2533	hypothetical prote
3	48.5	9.8	68	2 AD2173	hypothetical prote
4	48.5	9.8	96	2 G33730	Ig kappa chain V r
5	48	9.7	51	2 H82689	hypothetical prote
6	47.5	9.6	65	1 XT8G1	acrosin inhibitor
7	47.5	9.6	71	2 S20394	sperm-associated a
8	47.5	9.6	85	2 F86389	probable DNA-bind
9	47.5	9.6	92	2 A26395	T-cell receptor be
10	46	9.3	42	2 T07285	hypothetical prote
11	46	9.3	69	2 PH1080	Ig light chain V r
12	46	9.3	80	2 T01450	limonene cyclase h
13	46	9.3	87	2 A64429	hypothetical prote
14	45.5	9.2	69	2 E25647	hypothetical lc pr
15	44.5	9.0	58	2 AC1869	hypothetical prote
16	44.5	9.0	72	2 E43259	H+-transporting tw
17	44.5	9.0	88	2 S66287	sapacin B precursor
18	44.5	9.0	92	2 AB2021	hypothetical prote
19	44	8.9	68	2 C97874	degenerate transpo
20	44	8.9	72	2 T31010	hypothetical prote
21	43.5	8.8	60	2 S68769	short neurotoxin -
22	43.5	8.8	82	2 T04766	acclimation protei
23	43.5	8.8	84	2 T17637	hypothetical prote
24	43	8.7	66	2 H82783	hypothetical prote
25	43	8.7	67	2 B35063	hypothetical prote
26	43	8.7	71	2 T29378	lin-32 protein - C
27	43	8.7	92	2 AF1919	hypothetical prote
28	43	8.7	96	2 T45337	hypothetical prote
29	42.5	8.6	58	2 B43928	probable collagen

30	42.5	8.6	78	2 T09040	hypothetical prote
31	42.5	8.6	85	2 S58175	acyl carrier prote
32	42.5	8.6	85	2 E95976	hypothetical prote
33	42	8.5	50	1 INT02	insulin 2 - load#1
34	42	8.5	70	2 T07342	hypothetical prote
35	42	8.5	76	2 S09897	hypothetical prote
36	42	8.5	80	2 T27603	hypothetical prote
37	42	8.5	83	2 C96916	hypothetical prote
38	42	8.5	85	2 T14678	hypothetical prote
39	42	8.5	87	2 T68792	MHC class II histo
40	42	8.5	87	2 B23662	ribosomal protein
41	41.5	8.4	59	2 S10348	T-cell receptor be
42	41.5	8.4	77	1 A69093	hypothetical prote
43	41.5	8.4	81	2 B82560	hypothetical prote
44	41.5	8.4	85	2 E42825	kruppel-type zinc
45	41	8.2	55	2 AD1852	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S44828 F54F2.3 protein - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 23-Mar-2001

C/Accession: S44828

R/Anderson, K.

A/Description: Sequence of the C. elegans cosmid F54F2.

A/Reference number: S44817

A/Accession: S44828

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-34 <AND>

A/Cross-references: EMBL:L23645; NID:G386603; PID:G386609

C/Genetics:

A/Introns: 20/1

Query Match

Best Local Similarity 10.1%; Score 50; DB 2; Length 34;

Matches 10; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY

14 MCVNFKKQKRGCTVHREADFPWSICTADMSLL 47

DB 1 MKNVQMKRVKVVHCRVTSRQSLRNLCARLTFW 34

##### RESULT 2

AG2533 hypothetical protein asr7554 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120b

C/Species: Nostoc sp. PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C/Accession: AG2533

R/Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuch

Nakazaki, N.; Shindo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,

DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A/Reference number: AB1807; MUID:21595285; PMID:11759840

A/Accession: AG2533

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-74 <KIR>

A/Cross-references: GB:AP003602; PIDN:BA877197.1; PID:G17134639; GSPDB:GN00181

A/Experimental source: strain PCC 7120

C/Genetics:

A/Gene: asr7554

A/Genome: plasmid

Query Match

Best Local Similarity 9.9%; Score 49; DB 2; Length 74;

Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 25 GLCTVHREADF--FWSICTAD 43  
 DB 12 GICTFHWNADFFKFCQVCNPD 32

## RESULT 3

AD2173  
 hypochelical protein asr2939 [imported] - Nostoc sp. (strain PCC 7120)  
 C/Species: Nostoc sp. PCC 7120  
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C/Accession: AD2173  
 R/Kaneko, T.; Nakamura, Y.; Maki, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AD2173  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-68 <KDP>  
 A/Cross-references: GB:BA000019; PIDN:BA074638.1; PID:G17132033; GSPDB:GN00179  
 A/Experimental source: strain PCC 7120  
 A/Genetics: asr2939

Query Match 9.8%; Score 48.5; DB 2; Length 68;  
 Best Local Similarity 36.1%; Pred. No. 1.2e+02;  
 Matches 13; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

## RESULT 4

G33730  
 Ig kappa chain V region (23.32) - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 21-Jan-2000  
 C/Accession: G33730  
 R/Lawler, A.M.; Kearney, J.F.; Kuehl, M.; Gearhart, P.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 6744-6747, 1989  
 A/Title: Early rearrangements of genes encoding murine immunoglobulin kappa-chains, until A/Reference number: A33730; MUID:89367325; PMID:2505260  
 A/Accession: G33730  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-96 <LAW>  
 C/Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: heterodimer; immunoglobulin  
 F:16-90/Domain: immunoglobulin homology <IMM>

Query Match 9.8%; Score 48.5; DB 2; Length 96;  
 Best Local Similarity 29.5%; Pred. No. 1.9e+02;  
 Matches 13; Conservative 7; Mismatches 17; Indels 7; Gaps 1;

## RESULT 5

H82689  
 hypochelical protein Xr1364 [imported] - Xylella fastidiosa (strain 9a5c)  
 C/Species: Xylella fastidiosa  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C/Accession: H82689  
 R/Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences  
 Nature 406, 151-157, 2000  
 A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A/Reference number: AB2515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: H82689  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-51 <SIM>  
 A/Cross-references: GB:AE003968; GB:AE003649; NID:99106363; PIDN:AF84173.1; GSPDB:GN  
 A/Experimental source: strain 9a5c  
 R/Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.B.A.; Carraro, D.M.; Carre  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinanti, A.P.; Ferreira, A.J.S.  
 Submitted to Genbank, June 2000  
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; La  
 chado, M.A.; Madalza, A.M.B.N.; Madalza, H.M.F.; Marino, C.L.; Marques, M.V.; Martini  
 A/Authors: Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.F.M.; Miracca, E.C.; Miyaki, C.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.F.M.; Miracca, E.C.; Miyaki, C.  
 Rodrigues, V.; Rosa, A.J.; de M., de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava  
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A/Reference number: A59328  
 A/Contents: annotation  
 C/Genetics:  
 A/Genes: Xr1364

Query Match 9.7%; Score 48; DB 2; Length 51;  
 Best Local Similarity 28.1%; Pred. No. 1e+02;  
 Matches 16; Conservative 6; Mismatches 23; Indels 12; Gaps 2;

## RESULT 6

XTPG1  
 acrosin inhibitor A1 (pSTI type) - pig  
 C/Species: Sus scrofa domestica (domestic pig)  
 C/Date: 30-Nov-1980 #sequence\_revision 30-Nov-1980 #text\_change 16-Jul-1999  
 C/Accession: A01233  
 R/Teschke, H.; Kupfer, S.; Klausner, R.; Fink, E.; Fritz, H.  
 In: Provides of the Biological Fluids, Proc. 23rd Colloq., Peeters, H., ed., pp.255-26  
 A/Reference number: A94431  
 A/Accession: A01233  
 A/Molecule type: protein  
 A/Residues: 1-65 <TSC>  
 C/Comment: Acrosin inhibitor A1 is one of several closely related inhibitors found in  
 C/Superfamily: pancreatic secretory trypsin inhibitor; Kazal proteinase inhibitor hom  
 C/Keywords: glycoprotein; semen; serine proteinase inhibitor; sperm; testis  
 F:5-57/Domain: Kazal proteinase inhibitor homology <KPI>  
 F:7-39,17-36,25-57/Disulfide bonds: #status experimental  
 F:12,62/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:19/Inhibitory site: Arg (acrosin, trypsin) #status experimental

Query Match 9.6%; Score 47.5; DB 1; Length 65;  
 Best Local Similarity 26.9%; Pred. No. 1.5e+02;  
 Matches 14; Conservative 13; Mismatches 16; Indels 9; Gaps 3;

## RESULT 7

S20394  
 sperm-associated acrosin inhibitor - pig  
 C/Species: Sus scrofa domestica (domestic pig)  
 C/Date: 22-Nov-1993 #sequence\_revision 01-Sep-1995 #text\_change 16-Jul-1999  
 C/Accession: S20394; S20395; S68652  
 R/Donakova, V.; Calveite, J.U.; Mann, K.; Schaefer, W.; Schmid, E.R.; Toepfer-Petersen  
 FEBS Lett. 297, 147-150, 1992  
 A/Title: The complete primary structure of three isoforms of a boar sperm-associated  
 A/Reference number: S20394; MUID:92201370; PMID:1551420  
 A/Accession: S20394



A:Molecule type: protein  
A:Residues: 1-71 <JON>  
A:Experimental source: spermatozoa  
A:Accession: S20395  
A:Molecule type: protein  
A:Residues: 3-68 <UO2>  
A:Experimental source: spermatozoa  
A:Note: these proteins were not glycosylated  
R:Calvete, J.J.; Dostalova, Z.; Sanz, L.; Adermann, K.; Thole, H.H.; Toepfer-Petersen, E  
FMS Lett. 379, 207-211, 1996  
A:Title: Mapping the heparin-binding domain of boar spermadhesins.  
A:Reference number: S68648; PMID:9618456; PMID:8603690  
A:Accession: S68652  
A:Molecule type: protein  
A:Residues: 1-11, 'X', 13-18-21, 'X', 23-26 <CAL>  
C:Superfamily: pancreatic secretory trypsin inhibitor; Kazal proteinase inhibitor homolo  
C:Keywords: semen; serine proteinase inhibitor; sperm; testis  
F:1-71/Product: sperm-associated acrosin inhibitor, minor form #status experimental <MT  
F:1-68,70-71/Product: sperm-associated acrosin inhibitor, long form #status experimental  
F:3-68,70-71/Product: sperm-associated acrosin inhibitor, short form #status experimen  
F:10-62/Domains: Kazal proteinase inhibitor homology <KPI>  
F:12-44,22-41,30-62/Diulfide bonds: #status experimental  
F:50/inhibitory site: Arg (acrosin, trypsin) #status predicted

Query Match 9.6%; Score 47.5; DB 2; Length 71;  
Best Local Similarity 30.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 12; Mismatches 15; Indels 11; Gaps 4;

20 KQKRGELCTVHREAPFMSLCTADMSLD---EQSHSPSLVYQCLSQKLROER 69  
Db 5 KRRKEDCVVYSHLFF---CTRENDPICGNGKSYANPCIF---CSKLGRRNK 53

RESULT 8  
F86389  
probable DNA-binding protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C:Accession: F86389

R:Theologos, A.; Ecker, J.R.; Palm, C.U.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzar, L.  
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, U.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: F86389

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-85 <STO>  
A:Cross-references: GB:AE005172; NID:G11079506; PIDN:AG29217.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: transcription factor sqwa; serum response factor DNA-binding domain homol

Query Match 9.6%; Score 47.5; DB 2; Length 85;  
Best Local Similarity 28.3%; Pred. No. 2.1e+02;

Matches 13; Conservative 10; Mismatches 20; Indels 3; Gaps 1;

6 LLEVDGPMGNVPEFKQKRGELCTVHREAPFMSLCTADMSLLEQSH 51  
Db 8 LKRIENKINQVTFPSKRTGL--KKAQETISVLCAEVSILVPSH 50

RESULT 9  
A26395  
T-cell receptor beta chain V-D-J-C0 regions (NZM8) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jun-1988 #sequence\_revision 02-Jun-1988 #text\_change 30-May-1997

C:Accession: A26395  
R:Benlik, M.A.; Ioh, D.Y.  
Nature 322, 379-382, 1986  
A:Title: Alternative splicing of murine T-cell receptor beta-chain transcripts.  
A:Reference number: A26395; PMID:86284983; PMID:3488509  
A:Accession: A26395  
A:Molecule type: mRNA  
A:Residues: 1-92 <BEH>  
A:Cross-references: GB:M30880  
A:Note: alternative splicing gives rise to transcripts containing an additional exon be  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: T-cell receptor

Query Match 9.6%; Score 47.5; DB 2; Length 92;  
Best Local Similarity 30.0%; Pred. No. 2.3e+02;  
Matches 27; Conservative 6; Mismatches 26; Indels 31; Gaps 6;

5 SILEVDGPMGNVPEFKQKRG-----LCTVHREAPFMS-----LCT 41  
Db 7 SLME-DGAFKFD-RFKAMLNMSFSTLKIQTEPKDSAVLCASTRATEVFPFGKGLTV 64

42 ADMSLLEQSHSPSLVYQCLSQKLROERGT 71  
Db 65 VGLRLSYASHHS-SLTSQC-----RSECGT 88

RESULT 10  
T07285  
hypothetical protein 42b - Chlorella vulgaris chloroplast

C:Species: chloroplast Chlorella vulgaris  
C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000  
C:Accession: T07285

R:Makaseugi, T.; Nagai, T.; Kapoor, M.; Saito, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Naka  
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997

A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chl  
A:Reference number: Z15985; PMID:97303241; PMID:9158184

A:Accession: T07285

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-42 <MAX>

A:Cross-references: EMBL:AB001684; NID:G2224352; PIDN:BA57933.1; PID:G2224449

C:Genetics:

A:Genome: chloroplast

C:Keywords: chloroplast

Query Match 9.3%; Score 46; DB 2; Length 42;  
Best Local Similarity 28.2%; Pred. No. 1.4e+02;  
Matches 11; Conservative 9; Mismatches 9; Indels 10; Gaps 1;

48 EDSHSPSLVYQCLSQKLROERGTIPGSGITRESKDMFHS 86  
Db 8 KKSYYTGSIVLTIVDQLEEE-----EEDWFPS 36

RESULT 11  
PH1080

Ig light chain V region (clone 165.60) - mouse (fragment)

C:Species: Mus musculus (house mouse)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Jun-1996  
C:Accession: PH1080

R:Tillman, D.M.; Jou, N.T.; Hill, R.J.; Marion, T.N.  
J. Exp. Med. 176, 761-779, 1992

A:Title: Both IgM and IgG anti-DNA antibodies are the products of clonally selective B

A:Reference number: PH0971; PMID:9238144; PMID:1512540

A:Accession: PH1080

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-69 <TLV>

A:Experimental source: B cell, strain [NZB x NZW]F1

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: immunoglobulin

Query Match 9.3%; Score 46; DB 2; Length 69;

Best Local Similarity 28.0%; Pred. No. 2.5e+02;  
Matches 14; Conservative 6; Mismatches 18; Indels 12; Gaps 1;

QY 48 EQSHSPSLVLCISQKLRQERGTIPGSGI-----TESKDMHF 85  
DB 9 QKSHSPRLIKVAGSISGIPSRFSGSGGTFTLINSVEREDIGMWF 58

## RESULT 12

101450  
limonene cyclase homolog F2401.12 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001  
C/Accession: T01450  
R/Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Cor  
eologis, A.; Becker, J.R.  
Submitted to the EMBL Data Library, January 1998  
A/Description: Genomic sequence for Arabidopsis thaliana BAC F2401.  
A/Reference number: Z14211  
A/Accession: T01450  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-80 <SH>  
A/Cross-references: EMBL:AC003113; NID:G2689438; PID:G2781356; GSPDB:GN00059; ATSP:F2401  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: ATSP:F2401.12  
A/Map position: 1

Query Match 9.3%; Score 46; DB 2; Length 80;  
Best Local Similarity 26.5%; Pred. No. 3e+02;  
Matches 9; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 49 QSHSPSLVLCISQKLRQERGTIPGSGITRESKD 82  
DB 37 KSPFLPRLIKMIDKVRNRLTMSGCVLTKYKD 70

## RESULT 13

A64429  
hypothetical protein MJ1034 - Methanococcus jannaschii  
C/Species: Methanococcus jannaschii  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 29-Sep-1999  
C/Accession: A64429  
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A/Reference number: A64429; MIMD:96337999; PMID:8688087  
A/Accession: A64429  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-87 <BLU>  
A/Cross-references: GB:U67546; GB:U77117; NID:G1591687; PIDN:AA899038.1; PID:G1499877; T  
C/Genetics:  
A/Map position: conserved hypothetical protein MJ1034  
C/Superfamily: conserved hypothetical protein MJ1034

Query Match 9.3%; Score 46; DB 2; Length 87;  
Best Local Similarity 26.3%; Pred. No. 3.3e+02;  
Matches 10; Conservative 12; Mismatches 12; Indels 4; Gaps 2;

QY 7 LEVDGPAMKNVFKAKRGIL-CTHREADF---FWSLC 40  
DB 24 LAIEKPSLKDIKALKGLIEPKIYRDKRYPROHWEIC 61

## RESULT 14

E25647  
hypothetical 1c protein - phage PA2  
C/Species: phage PA2

A/Note: host Escherichia coli  
C/Date: 05-Jun-1998 #sequence\_revision 05-Jun-1998 #text\_change 08-Oct-1999  
C/Accession: E25647  
R/Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A/Title: Structure of the 1c and nmc outer membrane porin protein genes of lambdaoid  
A/Reference number: A25647; MIMD:86304457; PMID:3017988  
A/Accession: E25647  
A/Molecule type: DNA  
A/Residues: 1-69 <BLA>  
A/Cross-references: GB:J02580; NID:G215366; PIDN:AAA32302.1; PID:G215370

Query Match 9.2%; Score 45.5; DB 2; Length 69;  
Best Local Similarity 22.2%; Pred. No. 2.9e+02;  
Matches 12; Conservative 16; Mismatches 21; Indels 5; Gaps 2;

QY 44 MSLEQSHSP--SLVLCISQKLRQERGTIPGSGITRESKDMHFSISGILNDV 95  
DB 17 LEVNGTYENDVISAATBESFVIFVKQVDFV---INSRRIIFVEVIGCTYINI 67

## RESULT 15

AC1869  
hypothetical protein asr0500 [imported] - Nostoc sp. (strain PCC 7120)  
C/Species: Nostoc sp. PCC 7120  
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C/Accession: AC1869  
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A/Reference number: AB1807; MIMD:21595285; PMID:11759840  
A/Accession: AC1869  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-58 <KUR>  
A/Cross-references: GB:BA000019; PIDN:BA872458.1; PID:G17129845; GSPDB:GN00179  
A/Experimental source: strain PCC 7120  
C/Genetics:  
A/Gene: asr0500

Query Match 9.0%; Score 44.5; DB 2; Length 58;  
Best Local Similarity 52.0%; Pred. No. 3.1e+02;  
Matches 13; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 45 SLLEQ-SHSPSLVLCISQKLRQE 68  
DB 13 SLLEQSTYEOOEIIQVLSQKHSQ 37

Search completed: September 9, 2004, 06:13:59  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:12:29 ; Search time 10 Seconds  
(without alignments)  
499.873 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_253\_348  
Perfect score: 497  
Sequence: 1 LEDSLILEVDGPMKRVKVEFK.....ITKSKDHGFSGLGILLDVL 96

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 13876

Minimum DB seq length: 0  
Maximum DB seq length: 96

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.1\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	10.9	92	Y614_PASMU	O9CN31 Pasteurella
2	47.5	9.6	60	HXB1_BRARE	O4366 brachydanio
3	47.5	9.6	65	IACA_PIG	P00999 sus scrofa
4	46	9.3	87	SR19_METUA	O5840 methanococc
5	46	9.3	87	M8C_SCHUA	P1523 schistosoma
6	45	9.1	88	EF18_THEAC	O91KX1 thermoplasma
7	44.5	9.0	88	SABP_SARPE	P31529 sarcophaga
8	43.5	8.8	60	NX61_MICNI	P80548 micrurus ni
9	43	8.7	67	Y132_HALMA	P20572 halorcula
10	43	8.7	71	L132_CAREL	O10574 caenorhabdi
11	42	8.5	50	INS2_BATSP	P01338 batrachoidi
12	42	8.5	76	ULD1_HOMVA	P16773 human cytom
13	41.5	8.4	82	Y52A_METUA	P81309 methanococc
14	41.5	8.4	96	GATC_BACRA	O81269 bacillus an
15	41.5	8.4	96	GATC_BACRA	O81269 bacillus an
16	41	8.2	69	SLYX_PSEAE	O91429 pseudomonas
17	41	8.2	72	SLYX_PSEAE	O91429 pseudomonas
18	41	8.2	83	B1E2_STRAU	P22491 streptococ
19	41	8.2	81	RL27_STRAU	O82C86 streptococ
20	41	8.2	91	RACC_ECOLI	P18033 escherichia
21	41	8.2	95	ACP_SACER	P11830 saccharopol
22	40.5	8.1	95	Y930_METUA	O58397 methanococ
23	40.5	8.1	95	Y310_HAEIN	P43982 haemophilu
24	40	8.0	64	Y160_BPT4	P39251 bacterioph
25	40	8.0	64	Y134_ARCFU	O29931 archaeoglob
26	40	8.0	70	ATP5_VAIZE	O41898 zea mays (m
27	40	8.0	72	Y1IF_ECOLI	P32150 escherichia
28	40	8.0	89	RL27_CAUCR	O9a6b8 caulobacter
29	39.5	7.9	82	PER_CHRVI	P00208 chromobacter
30	39.5	7.9	91	YJF4_ECOLI	P33292 escherichia
31	39	7.8	51	INS_ACCOA	P01324 acemys cali
32	39	7.8	51	INS_ANSAN	P07454 anser anser
33	39	7.8	51	INS_BALBO	P01314 balaeopter

34	39	7.8	51	1	INS_BALPH	P01312 balaeopter
35	39	7.8	51	1	INS_CAMDR	P01320 camelus dro
36	39	7.8	51	1	INS_CAPHT	P01319 capra hircu
37	39	7.8	51	1	INS_DIDMA	P18109 didelphis m
38	39	7.8	51	1	INS_ELEWA	P01316 elephas max
39	39	7.8	51	1	INS_FELCA	P06306 felis silve
40	39	7.8	51	1	INS_TRASC	P11887 trachemys s
41	39	7.8	69	1	ATP5_ARATH	O96253 arabidopsis
42	39	7.8	69	1	ATP5_IPOBA	O06450 ipomoea bat
43	39	7.8	73	1	IF1_FUSNN	O815W2 fusobacteri
44	39	7.8	87	1	RI27_PROMA	O7VAN2 prochlorococ
45	39	7.8	89	1	HS3A_XENDA	P04120 xenopus lae

ALIGNMENTS

RESULT 1  
Y614\_PASMU STANDARD; PRT; 92 AA.  
AC O9CN31;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein PM0614.  
GN PM0614.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Pasteurella.  
OX NCBI\_Taxid=747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L., Paustian M.L., Whitlam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida PM70."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; AEO06098; AAK02698.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 92 AA; 10340 MW; 304338C8E07ECA377 CRC64;

Query Match 10.9%; Score 54; DB 1; Length 92;  
Best Local Similarity 26.3%; Pred. No. 19;  
Matches 26; Conservative 16; Mismatches 33; Indels 24; Gaps 4;

QY 2 EDSLILEVDGPMKRVKVEFKGLCTVHREADFFWSLCTAD-MSLEQSHSSP--SLY 57  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db 7 QEKELIMKMGDALRMKLYAOSKPKQVTPH-----FTFADSLMSIASINQVRSLSA 58  
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
QY 58 LQCSOKLROERGTIPGSGITSKMHRFSSLCILLDVL 96  
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db 59 VSLSKKL-----LSSKFLTYSALGMATLYL 85  
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 2  
HXB1\_BRARE STANDARD; PRT; 60 AA.  
ID HXB1\_BRARE  
AC O42366;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Homeobox protein Hox-B1 (Fragment).  
GN HXB1 OR HOXB1B.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Pando.
OX NCBI_TaxID=7953;
RN (1)
RP SEQUENCE FROM N.A.
RA Prince V.E., Moens C.B., Kimmel C.B., Ho R.K.;
RT "zebrafish hox genes: expression in the hindbrain region of wild-type
RT and mutants of the segmentation gene, valentino.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBO databases.
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC ACTS ON THE ANTERIOR BODY STRUCTURES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the Antp homeobox family. Labial subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y13950; CAA74288.1; -.
DR TRANSFAC; T03662; -.
DR ZFIN; ZDB-GENE-990415-101; hoxb1a.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD00010; HOMEBOX; 1.
DR PROSITE; PS00027; HOMEBOX_1, PARTIAL.
DR PROSITE; PS50071; HOMEBOX_2, 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
DR Transcription regulation.
RV NON_TER
FT DNA_BIND
FT SEQUENCE
SQ 60 AA; 6691 MW; IDBBB26F515FB62 CRC64;

Query Match
Best Local Similarity 9.6%; Score 47.5; DB 1; Length 60;
Matches 14; Conservative 8; Mismatches 18; Indels 3; Gaps 2;

OY 47 LEQSHSSPSLYLQ-CISQKLRQERGTGSGITTEKMHFSS 87
DB 3 LEINETGVKIMFQNRKQKREKEGLAPASS-TSSKLEDDQS 44

RESULT 3
IACA_PIG STANDARD; PRT; 65 AA.
AC P00999;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Seminal plasma acrosin inhibitor A1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN (1)
RP SEQUENCE, AND DISULFIDE BONDS.
RA Tescheche H., Kupfer S., Klausner R., Fink E., Fritz H.;
RT (In) Peeters H. (eds.);
RL Provides of the biological fluids, Proc. 23th colloquium, pp.255-266,
RL Pergamon Press, New York (1976).
CC -1- FUNCTION: Inhibits acrosin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Seminal plasma.
CC -1- SIMILARITY: Contains 1 Kazal-like domain.
DR PIR; A01233; XTPGL.
DR InterPro; IPR002350; kazal.

```

```

DR InterPro; IPR001239; Kazal_inhib.
DR Pfam; PF00050; Kazal_1.
DR PRINTS; PR00290; KAZALINHIBR.
DR SMART; SM00280; KAZAL; 1.
DR PROSITE; PS00282; KAZAL; 1.
RV Serine protease inhibitor; Semen; Glycoprotein.
FT DOMAIN 5 57 KAZAL-LIKE.
FT DISULFID 7 39
FT DISULFID 17 36
FT DISULFID 25 57
FT ACT_SITE 19 20 REACTIVE BOND.
FT CARBOHYD 12 12 O-LINKED (PROBABLE).
FT CARBOHYD 62 62 O-LINKED (PROBABLE).
SQ SEQUENCE 65 AA; 7595 MW; 8DB22707CBB50480 CRC64;

Query Match
Best Local Similarity 9.6%; Score 47.5; DB 1; Length 65;
Matches 14; Conservative 13; Mismatches 16; Indels 9; Gaps 3;

OY 22 QKRGCTVHREADFMSICTADMSL-----EQSHSSPELYQCUSQKLRQER 69
DB 2 RQPNQNVYRSHLFF---CTQMDPICGTGKNSYANPCIF--CSEKGLRNQK 48

RESULT 4
SR19_METUA STANDARD; PRT; 87 AA.
AC OS8440;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal recognition particle 19 kDa protein (SRP19).
GN SRP19 OR MJ1034.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=JAL-1 / DSK 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RT Science 273:1058-1073(1996).
RL 7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC similarity).
CC -1- FUNCTION: Signal-recognition-particle assembly, binds directly to
CC 7S RNA and mediates binding of the 54 kDa subunit of the SRP (By
CC similarity).
CC -1- SUBUNIT: Archaeal signal recognition particle consists of a 7S RNA
CC molecule of 300 nucleotides and two protein subunits: SRP4 and
CC SRP19 (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the SRP19 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67546; AAB99038.1; -.
DR PIR; A64429; A64429.
DR PDB; 1L9A; 28-JUN-02.
DR FDB; 1LNG; 10-JUL-02.

```

DR TIGR; M01034; -  
 DR HAMAP; MF\_00305; -; 1.  
 DR InterPro; IPR002778; SRP19.  
 DR Pfam; PF01922; SRP19; 1.  
 DR Prodom; PD006609; SRP19; 1.  
 DR Signal recognition particle;  
 KW Complete proteome; 3D-structure; RNA-binding; Ribonucleoprotein;  
 SO SEQUENCE 87 AA; 10352 MW; 0BF272957B69CEB8 CRC64;

Query Match 9.3%; Score 46; DB 1; Length 87;  
 Best Local Similarity 26.3%; Pred. No. 1.5e+02;  
 Matches 10; Conservative 12; Mismatches 12; Indels 4; Gaps 2;

QY 7 LEVDGPMKVEFKAKRL-CTVREADF--FMSLC 40  
 DB 24 LAIEKSLKDIKALKGLLEPKRYDKRYPRQHEIC 61

## RESULT 5

MA8C\_SCHUA STANDARD; PRT; 95 AA.

AC P13523;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Immunogenic miracidial antigen 8C (Fragment).

OS Schistosoma japonicum (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxId=6182;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Philippines;  
 RX MEDLINE=89261924; PubMed=2657419;  
 RA Scallan B.V., Bogitsh B.V., Carter C.E.;  
 RT "Characterization of a large gene family in Schistosoma japonicum  
 that encodes an immunogenic miracidial antigen.";

RL M01. Biochem. Parasitol. 33:105-112(1989).  
 CC -1- DEVELOPMENTAL STAGE: Miracidia.  
 CC -1- SIMILARITY: Belongs to the immunogenic miracidial antigen family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M26213; AAA29852.1; -  
 KM Antigen; Multigene family.  
 FT NON TER 1  
 SQ SEQUENCE 95 AA; 10556 MW; 43719D2B8BBA1A9 CRC64;

Query Match 9.3%; Score 46; DB 1; Length 95;  
 Best Local Similarity 30.6%; Pred. No. 1.7e+02;  
 Matches 15; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 41 TADMSLLEQSHSPSLYLQCLSOXLRQERGTIPSGGITESKDMHPSSTG 89  
 DB 34 TDDVDVDDSHSGSPQLQGGYGRNDHYGGNGYGGYTRPKQHGNGRG 82

RESULT 6  
 EF1B\_THEAC STANDARD; PRT; 88 AA.  
 AC O9KNT1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Elongation factor 1-beta (EF-1-beta).  
 GN EF1B OR TA0566.

OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 OX NCBI\_TaxId=2303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479972; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Wever H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 acidophilum.";

RL Nature 407:508-513(2000).  
 CC -1- FUNCTION: Promotes the exchange of GDP for GTP in EF-1-alpha/GDP,  
 thus allowing the regeneration of EF-1-alpha/GTP that could then  
 be used to form the ternary complex EF-1-alpha/GTP/ActRNA (by  
 similarity).

CC -1- SIMILARITY: Belongs to the EF-1-beta/EF-1-delta family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; A1445064; CAC11706.1; ALT\_INIT.  
 DR HSSP; O27734; IGH8.  
 DR HAMAP; MF\_00043; -; 1.  
 DR InterPro; IPR004542; aEF-1.beta.  
 DR InterPro; IPR001326; EF1\_BD.  
 DR Pfam; PF00736; EF1BD; 1.  
 DR TIGRPFAM; TIGR004489; aEF-1.beta; 1.  
 KW Elongation factor; Protein biosynthesis; Complete proteome.  
 SQ SEQUENCE 88 AA; 9749 MW; 573537C1BADE5DA7 CRC64;

Query Match 9.1%; Score 45; DB 1; Length 88;  
 Best Local Similarity 22.4%; Pred. No. 2e+02;  
 Matches 11; Conservative 13; Mismatches 17; Indels 8; Gaps 2;

QY 3 DSSLLEVDGPMKVEFKAKRL-CTVREADF--EADFPMSLCIADMSLLEQ 49  
 DB 16 DADIXSIDQVQRNIE-----GLCSINRMDVDVIGGLKXIKLEIIVQ 58

RESULT 7  
 SARP\_SARPE STANDARD; PRT; 88 AA.  
 ID P15229;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Sarcosin B precursor.

OS Sarcophaga peregrina (Flesh fly) (Boettcherisca peregrina).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
 OC Sarcophagidae; Sarcophaga.  
 OX NCBI\_TaxId=7386;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=95361929; PubMed=7635204;  
 RA Lee S.-R., Kurata S., Natori S.;  
 RT "Molecular cloning of cDNA for sarcosin B, an antibacterial protein of  
 Sarcophaga, and its detection in larval brain.";

RL FEBS Lett. 368:485-487(1995).  
 CC -1- SIMILARITY: Belongs to the EF-1-beta/EF-1-delta family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

RU Biochem. J. 291:275-279(1993).  
CC -1- FUNCTION: Sapecin, which are potent bactericidal proteins,  
CC are produced in response to injury. Sapecin B is cytotoxic to  
CC Gram-positive bacteria.  
CC  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC  
CC -1- TISSUE SPECIFICITY: Hemocytes and fat body.  
CC  
CC -1- INDUCTION: By injury to the larval cell wall.  
CC  
CC -1- SIMILARITY: Belongs to the arthropod defensin family. Subfamily 1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: S80571; AAB35004.1; -;  
DR PIR: S66287; S66287.  
DR InterPro: IPR001542; Defensin\_anpod.  
DR Pfam: PF01097; Arthro\_defensin; 1.  
DR PROSITE: PS00425; ARTHROPOD\_DEFENSINS; FALSE NEG.  
KW Insect immunity; Antibiotic; Cleavage on pair of basic residues;  
KW Signal.  
FT SIGNAL 1 24 POTENTIAL.  
FT PROPEP 25 34  
FT PEPTIDE 35 88 SAPECIN B.  
FT DISULFID 57 78  
FT DISULFID 64 84  
FT DISULFID 68 86  
SQ SEQUENCE 88 AA; 10041 MW; 9459A0AF3B0DE3D CRC64;  
  
Query Match 9.0%; Score 44.5; DB 1; Length 88;  
Best Local Similarity 33.3%; Pred. No. 2.3e+02;  
Matches 21; Conservative 6; Mismatches 19; Indels 17; Gaps 5;  
  
QY 6 LLEVDGPAMKN-VEFKAKRGKGLCTVHREADFWSLCTADMSLLEQSHSPSLVIG--CLS 62  
DB 35 LQELDGAALDEPAELNHRKRLTCEIDR-----SLC-----LL--HCKRLKYLAVYCSQ 80  
QY 63 QKL 65  
DB 81 QKV 83  
  
RESULT 8  
NXSL MITCN  
ID NXSL MITCN STANDARD; PRT; 60 AA.  
AC P80548;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Short neurotoxin 1 (Neurotoxin alpha).  
OS Micrurus nigrocinctus (Central American coral snake) (Gargantilla).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;  
OC Elapidae; Elapinae; Micrurus.  
OX NCBI\_Taxid=6635;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Venom;  
RX MEDLINE=9624643; PubMed=8665942;  
RA Rosso J. P., Vargas-Rosso O., Gutierrez J. M., Rochat H., Bougis P. E.,  
RT "Characterization of alpha-neurotoxin and phospholipase A2 activities  
RT from Micrurus Venoms. Determination of the amino acid sequence and  
RT receptor-binding ability of the major alpha-neurotoxin from Micrurus  
RT nigrocinctus nigrocinctus".  
RU Eur. J. Biochem. 238:231-239(1996).  
CC -1- FUNCTION: Produces peripheral paralysis by blocking neuromuscular  
CC transmission at the postsynaptic site. Binds to the nicotinic  
CC acetylcholine receptor.  
CC  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC  
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.

CC -1- SIMILARITY: Belongs to the snake toxin family.  
DR PIR: S68769; S68769.  
DR HSSP: P01426; INEA.  
DR InterPro: IPR003571; Snake\_toxin.  
DR Pfam: PF00087; toxin; 1.  
DR ProDom: PD000206; Snake\_toxin; 1.  
DR PROSITE: PS00272; SNAKE\_TOXIN; 1.  
KW Toxin; Neurotoxin; Postsynaptic neurotoxin.  
FT DISULFID 3 22 BY SIMILARITY.  
FT DISULFID 17 39 BY SIMILARITY.  
FT DISULFID 41 52 BY SIMILARITY.  
FT DISULFID 53 58 BY SIMILARITY.  
SQ SEQUENCE 60 AA; 6582 MW; 4E980F86F0279C CRC64;  
  
Query Match 8.8%; Score 43.5; DB 1; Length 60;  
Best Local Similarity 30.4%; Pred. No. 1.9e+02;  
Matches 14; Conservative 5; Mismatches 20; Indels 7; Gaps 2;  
  
QY 48 EQSHSPSLVLT-----QCLSQKLKRGKGLTPSSGT---TESKDMFSS 86  
DB 6 QQSOSPFTIKTCSQEQCYKKTWRDHRGRTISRGGCCPTVPGTHIS 51  
  
RESULT 9  
YL32 HALMA  
ID YL32 HALMA STANDARD; PRT; 67 AA.  
AC P20572;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 7.1 kDa protein in HMAL3 5' region (ORF2).  
OS Halosarcula marismortui (Halobacterium marismortui).  
OC Archaea; Euryarchaeota; Halobacteriales;  
OC Halobacteriaceae; Halosarcula.  
OX NCBI\_Taxid=2238;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90153945; PubMed=2406244;  
RA Arndt E., Kroemer W., Hatakeyama T.;  
RT "Organization and nucleotide sequence of a gene cluster coding for  
RT eight ribosomal proteins in the archaeobacterium Halobacterium  
RT marismortui".  
RU J. Biol. Chem. 265:3034-3039(1990).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J05232; AAB6858.1; -;  
DR PIR: B35063; B35063.  
DR InterPro: IPR003750; DUF171.  
DR Pfam: PF02598; DUF171; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 67 AA; 7080 MW; 30B7DD259BB42CAE CRC64;  
  
Query Match 8.7%; Score 43; DB 1; Length 67;  
Best Local Similarity 25.6%; Pred. No. 2.5e+02;  
Matches 21; Conservative 13; Mismatches 24; Indels 24; Gaps 4;  
  
QY 17 VEFKQKRGKGLCTVHREADFWSLCTADMSLLEQSHSP--SLYLQCLSQKLKRGKGLTP 73  
DB 3 VAFGAPERGLPSILDVA-----PDVAGDQCSDBEGRDLTL-----NTV 43  
QY 74 --GSGITBSKMKMHPSSLCGILL 93  
DB 44 NQSEVVRTEALFASLTCLTL 65  
  
RESULT 10

L132\_CABEL  
 ID L132\_CABEL STANDARD; PRT; 71 AA.  
 AC 010574;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Abnormal cell lineage protein 32.  
 GN LIN-32 OR T14F9.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND MUTAGENESIS OF  
 RP GUU-10 AND LEU-24.  
 RC STRAIN=Briscoll N2;  
 RX MEDLINE=95098129; PubMed=7800042;  
 RA Zhao C., Emmons S.W.;  
 RT "A transcription factor controlling development of peripheral sense  
 RT organs in C. elegans.";  
 RL Nature 373:74-78(1995).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscoll N2;  
 RA Miller N., Langston Y.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Essential for the specification of the neuroblast  
 CC cell fate in the development of peripheral sense organs.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U15418; AAA67360.1; -;  
 DR EMBL; U50199; AAA91264.1; -;  
 DR PIR; T29378; T29378.  
 DR WormPep; T14F9.5; CE07501.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS50888; HLH; 1.  
 KM Neurogenesis; Differentiation; Developmental protein; DNA-binding;  
 KM Nuclear protein.  
 FT DOMAIN 11 54 HELIX-LOOP-HELIX MOTIF.  
 FT MUTAGEN 10 10 TAIL TOUCH INSENSITIVITY.  
 FT MUTAGEN 24 24 L->F: IN LIN-32(E1926); 81% RAY LOSS.  
 FT SEQUENCE 71 AA; 8318 MW; 3CF38A61B8E7CB7F CRC64;  
 SQ  
 Query Match 8.7%; Score 43; DB 1; Length 71;  
 Best local similarity 58.3%; Pred. No. 2.7e+02;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 57 YVQCLSQKRLQOE 68  
 DB 49 YIECLSQRLKOD 60  
 RESULT 11  
 INS2\_BATSP STANDARD; PRT; 50 AA.  
 AC P01338;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Insulin 2.  
 OS Batrachoididae sp. (Toadfish).  
 CC -----

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphna; Paracanthopterygii; Batrachoididae.  
 NCBI\_TaxID=8066;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=66160119; PubMed=5949593;  
 RA Smith L.F.;  
 RT "Species variation in the amino acid sequence of insulin.";  
 RL Am. J. Med. 40:662-666(1966).  
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It  
 CC increases cell permeability to monosaccharides, amino acids and  
 CC fatty acids. It accelerates glycolysis, the pentose phosphate  
 CC cycle, and glycogen synthesis in liver.  
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two  
 CC disulfide bonds.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to the insulin family.  
 DR PIR; A01605; INT02.  
 DR HSSP; P01308; IAI0.  
 DR InterPro; IPR004825; Ins/IGF/relax.  
 DR PRINTS; PR00277; INSULINB.  
 DR SMART; SM00078; IIGF; 1.  
 DR PROSITE; PS00262; INSULIN; 1.  
 KM Insulin family; Hormone; Glucose metabolism.  
 FT CHAIN 1 29 INSULIN B CHAIN.  
 FT NON\_CONS 29 30  
 FT CHAIN 30 50 INSULIN A CHAIN.  
 FT DISULFID 8 36 INTERCHAIN.  
 FT DISULFID 20 49 INTERCHAIN.  
 FT DISULFID 35 40  
 SQ SEQUENCE 50 AA; 5652 MW; 903B8ACBD62137C CRC64;  
 QY 50 SHSPSLYVQCISQKLRQGRITPGSGITE 79  
 DB 10 SHLVDAIYVVC-----GDRGFYNSGIVE 33  
 RESULT 12  
 U1D1\_HOMVA STANDARD; PRT; 76 AA.  
 AC P16773;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Hypothetical protein U131.  
 GN U131.  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 NCBI\_TaxID=10360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90269039; PubMed=2161319;  
 RA Chee W.S., Bankier A.T., Beck S., Bohni R., Brown C.W., Cerny R.,  
 RA Hornell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,  
 RA Predlie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
 RT "Analysis of the protein-coding content of the sequence of human  
 RT cytomegalovirus strain AD169.";  
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



DR EMBL, X17403, CAA35294.1, --  
KW PIR, S09897, S09897.  
DR Hypothetical protein.  
FT CARBOHYD 70 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 76 AA; 8243 MW; 25CE1E15B5F8D1 CRC64;  
Query Match 8.5%; Score 42; DB 1; Length 76;  
Best Local Similarity 26.7%; Pred. No. 3.9e+02;  
Matches 16; Conservative 8; Mismatches 18; Indels 18; Gaps 3;  
OY 26 LCTVHREADPFWSLCTADMSILPQSHSSPSLYOCLSQKROERGTPGSG-----ITE 79  
DB 1 MCMVSHKAFLLS-----QHAAYSQVAVCLSS--VRGAGSVAPNAGKKTITTE 48  
RESULT 13  
Y52A\_METUA STANDARD; PRT; 82 AA.  
ID Y52A\_METUA  
AC P81309;  
DT 15-JUL-1998 (Rel. 36, Last Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M0526.1.  
GN M0526.1.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
CX NCBI\_TaxId=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; Pubmed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kesteven A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,  
RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurt M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii";  
RL Science 273:1058-1073(1996).  
CC -1- SIMILARITY: STRONG, TO M.THERMOAUTOTROPHICUM MTH366.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U67502; AAB98527.1; --  
DR TIGR: M0526.1; --  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 82 AA; 9169 MW; 4ABFA148BDE787EA CRC64;  
Query Match 8.4%; Score 41.5; DB 1; Length 82;  
Best Local Similarity 46.4%; Pred. No. 4.8e+02;  
Matches 13; Conservative 3; Mismatches 11; Indels 1; Gaps 1;  
OY 35 FFWSLCTADMSILPQSHSSPSLYOCLQS 62  
DB 11 FIALMTARI-FLERSARKLYLCLLS 37  
RESULT 14  
GATC\_BACCA STANDARD; PRT; 96 AA.  
ID GATC\_BACCA  
AC O81Z89;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Aspartyl-glutamyl-tRNA(Asn/Gln) amidotransferase subunit C  
DE (EC 6.3.5.-) (Asp/Glu-ADP subunit C).  
GN GATC OR BA0320.  
OS Bacillus anthracis (strain Ames).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
CX NCBI\_TaxId=198094;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=2608414; Pubmed=12721629;  
RX Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,  
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
RA Holtapple E.K., Oksard O.A., Helgason E., Ristone J., Wu M.,  
RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gwinn M.,  
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Niemeyer W.C.,  
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,  
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,  
RA Fraser C.M.;  
RT "The genome sequence of Bacillus anthracis Ames and comparison to  
RT closely related bacteria";  
RL Nature 423:81-86(2003).  
CC -1- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)  
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-  
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both  
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction  
CC takes place in the presence of glutamine and ATP through an  
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.  
CC -1- CATALYTIC ACTIVITY: ATP + L-asparaginyl-tRNA(Asn) + L-glutamine = ADP  
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.  
CC -1- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).  
CC -1- SIMILARITY: Belongs to the gacC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE017025; AAP24353.1; --  
DR TIGR: BA0320; --  
DR HAMAP: MF\_00122; -- 1.  
DR InterPro: IPR004415; GATC.  
DR InterPro: IPR003837; Gln-tRNA-Gln.  
DR Pfam: PF02686; Gln-tRNA-Gln; 1.  
DR TRIPFAMS: TIGR00135; gacC; 1.  
KM Protein biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE 96 AA; 10866 MW; 7A33F855139492F CRC64;  
Query Match 8.4%; Score 41.5; DB 1; Length 96;  
Best Local Similarity 19.7%; Pred. No. 5.9e+02;  
Matches 14; Conservative 18; Mismatches 30; Indels 9; Gaps 2;  
OY 13 AMKNEVEFKRGKGLCTVHREADPFWSLCTADMSILPQSHS-----SPSYLCTLSQKLR 66  
DB 5 SVENKHYAHLARLAIITQGEAEKFKQKQDAIVTFAEQUNELDTTVDVPTTHVLKRYMR 64  
OY 67 QERGITPGSGI 77  
DB 65 ED---VPEKGL 72  
RESULT 15  
GATC\_BACCR STANDARD; PRT; 96 AA.  
ID GATC\_BACCR  
AC O81N4;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)



DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Aspartyl-glutamyl-tRNA(Asn/Gln) amidotransferase subunit C  
DE (EC 6.3.5.-) (Asp/Glu-ADT subunit C).  
GN GATC OR BC0350.  
OS *Bacillus cereus* (strain ATCC 14579 / DSM 31).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
OX NCBI\_TaxID=226900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608415; PubMed=12721630;  
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,  
RA Kapralov V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,  
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
RA Overbeek R., Kyrpides N.;  
RT "Genome sequence of *Bacillus cereus* and comparative analysis with  
RT *Bacillus anthracis*.";  
RL Nature 423:87-91(2003).  
CC -1- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)  
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-  
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both  
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction  
CC takes place in the presence of glutamine and ATP through an  
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.  
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP  
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.  
CC -1- SUBUNIT: Heterotrimer of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits (By similarity).  
CC -1- SIMILARITY: Belongs to the gacC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AE016999; AAP07390.1; -.  
DR HAMAP; MF\_00122; -; 1.  
DR InterPro; IPR004415; GacC.  
DR InterPro; IPR003837; Glu-tRNA(Gln).  
DR Pfam; PF02686; Glu-tRNA(Gln); 1.  
DR TIGRFAMs; TIGR00135; gacC; 1.  
KW Protein biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE 96 AA; 10866 MW; 7A33F8551B39492F CRC64;

Query Match 8 4%; Score 41.5; DB 1; Length 96;  
Best Local Similarity 19.7%; Pred. No. 5.9e+02;  
Matches 14; Conservative 18; Mismatches 30; Indels 9; Gaps 2;  
QY 13 AMKNVEFKAKRGSLCTVREADPFWSLCTADMSLLEQSHS-----SPSLYLQICISQKLR 66  
DB 5 SVENYKHVHLRLAFLITDQEAKEFKQXDALVTFAEQNLNEDLTVDVKKPTHTVLTAKNVR 64  
QY 67 QERGITPGSGI 77  
DB 65 ED--VPEKGL 72

Search completed: September 9, 2004, 06:15:13  
Job time : 12 secs

This page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:10:00 ; Search time 39 Seconds

(without alignments)  
776,660 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_253\_348

Sequence: 1 LEDSLLEVDGPMANVEFK.....ITSKDMFSSLCILLDLV 96

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 157800

Minimum DB seq length: 0  
Maximum DB seq length: 96

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	11.7	62	12 Q8VAX8	Q8VAX8 white spot
2	57.5	11.6	91	10 Q8W459	Q8W459 arabidopsis
3	55.5	11.2	93	16 Q8EKC4	Q8EKC4 shewanella
4	55	11.1	78	16 Q7U750	Q7U750 synchococc
5	53	10.7	65	2 Q9ZG19	Q9ZG19 chlamydia t
6	50.5	10.2	64	10 Q42305	Q42305 arabidopsis
7	50	10.1	90	6 Q7YR94	Q7YR94 bos taurus
8	49.5	10.0	92	17 Q8TQY7	Q8TQY7 metanosc
9	49.5	10.0	91	12 Q7T6T2	Q7T6T2 canine coro
10	49.5	10.0	94	6 Q9BFC4	Q9BFC4 tepirus ind
11	49.5	10.0	96	10 Q8AN66	Q8AN66 oryza sativ
12	49	9.9	74	16 Q8ZS76	Q8ZS76 anabaena sp
13	48.5	9.9	92	16 Q8DUM3	Q8DUM3 streptococc
14	48.5	9.8	68	16 Q8YS20	Q8YS20 anabaena sp
15	48.5	9.8	74	6 Q95LE1	Q95LE1 canis fami
16	48.5	9.8	75	4 Q9UBA7	Q9UBA7 homo sapien

17	48.5	9.8	87	12 Q9B672	Q9B672 simian cyto
18	48.5	9.8	94	6 Q9BF05	Q9BF05 ceratotheri
19	48.5	9.8	96	16 Q89YD2	Q89YD2 bradyrhizob
20	48	9.7	51	16 Q9PD15	Q9PD15 xylella fas
21	47.5	9.6	59	3 Q9B6C3	Q9B6C3 coprinus ci
22	47.5	9.6	83	2 Q8KX21	Q8KX21 proteus vul
23	47.5	9.6	85	10 Q9C576	Q9C576 arabidopsis
24	47	9.5	58	4 Q9P148	Q9P148 homo sapien
25	47	9.5	64	16 Q7VA48	Q7VA48 prochloroco
26	47	9.5	84	5 Q8IPU6	Q8IPU6 dirosophila
27	46.5	9.4	58	4 Q9BS37	Q9BS37 homo sapien
28	46.5	9.4	61	4 Q96C06	Q96C06 homo sapien
29	46.5	9.4	76	16 Q89FW8	Q89FW8 bradyrhizob
30	46.5	9.4	90	16 Q920T0	Q920T0 rhizobium m
31	46.5	9.4	96	4 Q96CY8	Q96CY8 homo sapien
32	46	9.3	56	16 Q83BU1	Q83BU1 coxiella bu
33	46	9.3	66	6 Q81005	Q81005 macaca fasc
34	46	9.3	74	1 Q9V2U6	Q9V2U6 methanococc
35	46	9.3	83	10 Q8RV72	Q8RV72 arabidopsis
36	46	9.3	85	15 Q86579	Q86579 human immun
37	46	9.3	86	11 Q9QX28	Q9QX28 mus musculu
38	46	9.3	91	15 Q70910	Q70910 human immun
39	46	9.3	94	6 Q9BE29	Q9BE29 ursus arcto
40	46	9.3	95	6 Q9BF37	Q9BF37 choleopus d
41	46	9.3	95	6 Q9BF38	Q9BF38 choleopus h
42	46	9.3	95	6 Q9BF02	Q9BF02 leopards p
43	45.5	9.2	69	9 Q35866	Q35866 bacterioph
44	45.5	9.2	83	16 Q99QD6	Q99QD6 streptomyce
45	45.5	9.2	95	10 Q9S288	Q9S288 oryza sativ

#### ALIGNMENTS

RESULT 1  
Q8VAX8 PRELIMINARY; PRT; 62 AA.  
AC Q8VAX8;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE W5V241 (W5SV297).  
OS White spot syndrome virus (WSSV).  
OC Viruses; dsDNA viruses, no RNA stage; Nimaviridae.  
OX NCBI\_Taxid=92652;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21548311; PubMed=11689662;  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RT "Complete genome sequence of the shrimp white spot bacilliform  
RT virus.";  
RL J. Virol. 75:11811-11820 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RL Submitted (Dec-2000) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RX MEDLINE=20517548; PubMed=11062040;  
RA Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,  
RT Wang C.H., Lin U.Y., Kou G.H., Lo C.F.;  
RT "Identification and characterization of a shrimp white spot syndrome  
RT virus (WSSV) gene that encodes a novel chimeric polypeptide of  
RT cellular-type thymidine kinase and thymidylate kinase.";  
RL Virology 277:100-110 (2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RX MEDLINE=21844071; PubMed=11853398;  
RA Chen L.L., Leu U.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,  
RA Lo C.F., Kou G.H.;  
RT "Identification of a nucleocapsid protein (VP35) gene of shrimp white

RT spot syndrome virus and characterization of the motif important for  
targeting VP35 to the nuclei of transfected insect cells." ;  
RT Virology 293:44-53(2002).

RP SEQUENCE FROM N.A.  
RC STRAIN=Taiwan;  
RA Lo C.-F., Kou G.-H.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF332093; AL33245.1; -  
DR EMBL: AF440570; AL169165.1; -  
SQ SEQUENCE 62 AA; 6698 MW; 9B8CCD06C5367855 CRC64;

Query Match 11.7%; Score 58; DB 12; Length 62;  
Best Local Similarity 31.8%; Pred. No. 17;  
Matches 21; Conservative 5; Mismatches 18; Indels 22; Gaps 4;

QY 21 ACRGCTVREADFFMSICTADMSLLBQSHSPSLYLQCLT-----QKLRQEGTI-PG 74  
DB 8 ADRKGITTKMS-----WHLIVADIS-----RCCSLFTTTAKTKQSSGTTSPG 51

QY 75 SGITGS 80  
DB 52 SVVDPS 57

## RESULT 2

Q8W459 PRELIMINARY; PRT; 91 AA.

AC Q8W459;  
DT 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Camrinoi P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinzaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY062831; AL32509.1; -  
KM Hypothetical protein.  
FT NON TER 1

SQ SEQUENCE 91 AA; 10166 MW; 23B136FF2CDD3AF1 CRC64;

Query Match 11.6%; Score 57.5; DB 10; Length 91;  
Best Local Similarity 32.2%; Pred. No. 31;  
Matches 19; Conservative 6; Mismatches 19; Indels 15; Gaps 2;

QY 50 SHSPSLYLQCLT-----QKLRQEGT-----TIGSGITSKDMHPSLQCLIL 93  
DB 22 SNSSSLFLTCLSHVLIITHTKTRDRGVKSTGATLTGFAVVDVMTFPGGLVRLM 80

## RESULT 3

Q8ECX4 PRELIMINARY; PRT; 93 AA.

AC Q8ECX4;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN SO2998.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella

OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NR-1;

RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Melne B., Clayton R.A.,  
RA Meyer T., Tsaplin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Imbrailo M., Lee K., Berry K., Lee C.,  
RA Mueller U., Khouri H., Gill U., Utterback T.R., McDonald L.A.,  
RA Feldlyum T.V., Smith H.O., Venter J.C., Nelson K.R., Fraser C.M.,  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL: AE015737; AAN56010.1; -  
DR TIGR: SO2998; -

KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 93 AA; 10415 MW; 90E978B046E89693 CRC64;

Query Match 11.2%; Score 55.5; DB 16; Length 93;  
Best Local Similarity 30.4%; Pred. No. 56;  
Matches 14; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY 49 QSHSPSLYLQCLTQKLRQEGTIPGSGITSKDMHPSLQCLILD 94  
DB 5 QAHATSDALVMTLSLELQER-QAKENGIAAYCDRIAEINQLITD 49

## RESULT 4

Q7U750 PRELIMINARY; PRT; 78 AA.

AC Q7U750;  
DT 01-OCT-2003 (TRENBLrel. 25, Created)  
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical.

GN SYNW1136.  
OS Synechococcus sp. (strain WH8102).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
OX NCBI\_TaxID=84588;

RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=22825697; PubMed=12917641;  
RA Palenik B., Brahmsha B., Larimer F.W., Land M., Hauser L., Chain P.,  
RA Lamerdin U., Regala W., Allen E.E., McCarron C., Paulsen I.,  
RA Dufresne A., Patensky F., Webb E.A., Metelbury J.,  
RT "The genome of a motile marine Synechococcus".  
RL Nature 424:1037-1042(2003).  
DR EMBL: BX569692; CAB07651.1; -  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 78 AA; 8419 MW; BC7F745AB7D6F8BA CRC64;

Query Match 11.1%; Score 55; DB 16; Length 78;  
Best Local Similarity 30.2%; Pred. No. 53;  
Matches 13; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 37 WSLCTADMSLLBQSHSPSLYLQCLTQKLRQ-----ERGITPSGI 77  
DB 28 WALC-----EGRELTSVLVQCVELISIRLXKNGSIPSAI 62

## RESULT 5

Q9ZG19 PRELIMINARY; PRT; 65 AA.

AC Q9ZG19;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Dihydrofolate reductase/thymidylate synthase (Fragment).  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.

```

RT "Assessment of the gene content of the chromosomal regions flanking
RT bovine DGAT1."
RL Genomics 0:0-0(2003).
DR EMBL: AF518964; CAD56801.1; -.
KM Hypothetical protein..
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 90 AA; 9811 MW; 601E466AF0B529ED CRC64;
OY Query Match 10.1%; Score 50; DB 6; Length 90;
Db Best Local Similarity 25.4%; Pred. No. 2,6e+02;
Matches 16; Conservative 8; Mismatches 19; Indels 20; Gaps 2;
OY 24 RGLCTVHEADPFW-----SLTAMSLLEQSHSPSLYIQCSQK 64
Db 27 RELLARFVADLFMMSECFSTKDPALRLSLVGSVCMTQALCSSHSS-SFLSRKAE 85
OY 65 LRQ 67
Db 86 VAO 88
RESULT 8
O8TOY7 PRELIMINARY; PRT; 92 AA.
AC O8TOY7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted protein.
GN MA1399.
OS Methosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
CX NCBI_TaxID=2214;
EN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Galvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor J., Strange-Thomson N., Dabellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrelli A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grihane D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett W., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:533-542(2002).
DR EMBL: AE010809; AAM04815.1; -.
SQ Complete proteome.
OY SEQUENCE 92 AA; 10709 MW; D2D387A2FA836AC2 CRC64;
OY Query Match 10.1%; Score 50; DB 17; Length 92;
Db Best Local Similarity 31.8%; Pred. No. 2,7e+02;
Matches 14; Conservative 9; Mismatches 13; Indels 8; Gaps 2;
OY 34 DFFW-----SLCTADMSLLEQSHSPSL--YIQCSQKLRQR 69
Db 42 DFFWLSVVGISICTSLNLCGNLCAQFFLGSSTVGLTEKRRTRK 85
RESULT 9
O7T6T2 PRELIMINARY; PRT; 71 AA.
AC O7T6T2;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

```

DE Non-structural protein 3a.  
GN NSP 3A.  
OS Canine coronavirus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
CC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BG10;  
RA Sanchez-Morgado J.M.;  
RT "Molecular Characterization of Virulent Canine Coronavirus BGF  
Strain."  
RT Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY342160; AA017221.1; -  
SQ SEQUENCE 71 AA; 7706 MW; F89CAA3CC72D134D CRC64;

Query Match 10.0%; Score 49.5; DB 12; Length 71;  
Best Local Similarity 26.6%; Pred. No. 2.3e+02;  
Matches 17; Conservative 18; Mismatches 24; Indels 5; Gaps 2;  
CY 34 DFFMSLCTADMSLIBQ---SHSGPSLYLQCLSQKROERGTPGSGITSKOMHPSLGC 90  
DB 2 DIVASIDTSDVAIDEDPCAHFATLVKEFKGKLVLCIGF--GDTLEAKDKAKAVKGC 59  
CY 91 ILDD 94  
DB 60 SIIE 63

RESULT 10  
O9BF04 PRELIMINARY; PRT; 94 AA.  
ID O9BF04  
AC O9BF04  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Preprolactin (Fragment).  
GN PNOG.  
OS Taurus indicus (Asiatic tapir) (Malayan tapir).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.  
CX NCBI\_TaxID=9602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21082082; PubMed=11214319;  
RA Murphy W.J., Elstirk E., Johnson W.E., Zhang Y.P., Ryder O.A.,  
O'Brien S.J.;  
RT "Molecular phylogenetics and the origins of placental mammals."  
RL Nature 409:614-618(2001).  
DR EMBL; AY011857; AG38306.1; -  
DR GO; GO:0007218; P:neuropeptide signaling pathway; IEA.  
DR InterPro; IPR006023; Opioid neuropep.  
DR Pfam; PF01160; Opiods\_neuropep; 2.  
FT NON\_TER 1  
FT NON\_TER 94  
SQ SEQUENCE 94 AA; 10441 MW; 8FF1A35616592845 CRC64;

Query Match 10.0%; Score 49.5; DB 6; Length 94;  
Best Local Similarity 24.6%; Pred. No. 3.1e+02;  
Matches 17; Conservative 12; Mismatches 19; Indels 21; Gaps 4;  
CY 36 FMSLCT-----ADMSLIEQSSPSLYLQCLSQ-----KIRO-----ERGTPG--- 74  
DB 5 FMPPTCTVVTGRSGROLSSADSEHVAALYOPASSTHUKRMFRHSLPQVQKGTPEGMDE 64  
CY 75 SGITSEKDM 83  
DB 65 AGEVEQKQL 73

RESULT 11  
O84NC6 PRELIMINARY; PRT; 96 AA

AC O84NC6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Transcription factor MADS23  
OS Oryza sativa (Japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Eriatoidae; Oryzae; Oryza.  
CX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Dongjin;  
RA Lee S., Kim J., An G.;  
RT "Identification and mutant isolation of rice MADS box genes."  
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AY177694; AA047704.1; -  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; P:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR002100; TF\_MADSbox.  
DR Pfam; PF00319; SRF-TE; 1.  
DR PRINTS; PR00404; MADSOMAIN.  
DR SMART; SMO0432; MADS; 1.  
DR PROSITE; PS00350; MADS\_BOX\_1; 1.  
DR PROSITE; PS00066; MADS\_BOX\_2; 1.  
SQ SEQUENCE 96 AA; 10658 MW; A4C219A3EEFE19B6 CRC64;

Query Match 10.0%; Score 49.5; DB 10; Length 96;  
Best Local Similarity 31.1%; Pred. No. 3.2e+02;  
Matches 14; Conservative 7; Mismatches 21; Indels 3; Gaps 1;

CY 9 VGPAMKVEFKAQKRGCTVHREADFFMSLCTADMSLIESSHS 53  
DB 11 INMATSROYTFKRSRSGLFKARELST---LQDAEVLGVVSVTS 52  
RESULT 12  
O8ZSF6 PRELIMINARY; PRT; 74 AA.  
ID O8ZSF6  
AC O8ZSF6  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein Asr7554.  
GN ASR7554.  
OS Anabaena sp. (strain PCC 7120).  
OC Plantae; PCC7120beta.  
CC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
CX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Saemoto S.,  
Ratland A., Itiguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120."  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003602; BAB77197.1; -  
DR FTR; AG2533; AG2533.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
KW plasmid; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 74 AA; 8875 MW; 01580F056A9E9262 CRC64;

Query Match 9.9%; Score 49; DB 16; Length 74;  
Best Local Similarity 47.6%; Pred. No. 2.8e+02;  
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

CY 25 GLCTVHREAD--FMSLCTAD 43

Db 12 GICTFHMANDFYKFCQVCNFD 32

RESULT 13

```
Q8DUM3 PRELIMINARY; PRT; 92 AA.
ID Q8DUM3
AC Q8DUM3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN SMU.896.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RA MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., Moshan W.M., McLaughlin R.B., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE014931; AAN58609.1; -.
DR InterPro; IPR004386; Cons_hypoch53.
DR Pfam; PF03890; DUF332.1_hypoch53.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 10723 MW; 3467121C16CE0E8E CRC64;
```

Query Match 9.8%; Score 48.5; DB 16; Length 68;  
Best Local Similarity 36.1%; Pred. No. 2.9e+02;  
Matches 13; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

QY 50 SHSPSLYLOCLSQKROERTGTPSGITES-KDMH 84  
Db 15 STAKPWLIVQLRKKGQEHRTITPANYAEHLADIH 50

RESULT 15

```
Q95LE1 PRELIMINARY; PRT; 74 AA.
ID Q95LE1
AC Q95LE1;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Interleukin-2 (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Chamizo C., Rubio J.M., Moreno J., Alvar J.;
RT "Semi-quantification of canine cytokine expression by one tube RT-
RT PCR.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF32867; AAL26925.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005134; F:interleukin-2 receptor binding; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR00779; Interleukin-2.
DR Pfam; PF00715; IL2; 1.
DR PRINTS; PR00265; INTERLEUKIN2.
DR ProDom; PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.
DR PROSITE; PS00424; INTERLEUKIN_2; 1.
FT NON_TER 1
FT NON_TER 74
SQ SEQUENCE 74 AA; 8529 MW; 9E9D560F0BA4DF18 CRC64;
```

Query Match 9.8%; Score 48.5; DB 6; Length 74;  
Best Local Similarity 30.0%; Pred. No. 3.2e+02;  
Matches 9; Conservative 11; Mismatches 7; Indels 3; Gaps 1;

QY 57 YLOCLSQKROERTGTPSGITES-KDMHFS 86  
Db 17 HLOCLAEELKUNLEAYL---GLPQSKNVAHLT 43

Search completed: September 9, 2004, 06:13:09  
Job time : 42 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:11:27 ; Search time 54 Seconds

(without alignments)  
502.307 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_253\_348  
Perfect score: 497  
Sequence: 1 LEDSGLEVDGPAMKNVEFK.....ITKSDNHFFSLGILLDVL 96

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 878024

Minimum DB seq length: 0  
Maximum DB seq length: 96

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	71.0	76	4	AAI15359 Peptide #
2	353	71.0	76	4	ABR34356 Peptide #
3	353	71.0	76	4	AAW27837 Peptide #
4	353	71.0	76	4	ABR29199 Peptide #
5	353	71.0	76	4	ABR19775 Protein #
6	353	71.0	76	4	AAW67546 Human bon
7	353	71.0	76	4	AAW55153 Human bra
8	353	71.0	76	4	ABG49189 Human liv
9	353	71.0	76	4	AAW03118 Peptide #
10	353	71.0	76	5	ABG37740 Human pep
11	60	12.1	92	3	AAW33250 Arabidops
12	58	11.7	71	3	AAW60552 Arabidops
13	58	11.7	72	3	AAW56185 Arabidops
14	57.5	11.6	64	4	AAW90887 Human imm
15	57.5	11.6	78	3	AAW61267 Arabidops
16	57	11.5	89	2	AAW82628 Ehrlichia
17	57	11.5	89	3	AAW78548 Ehrlichia
18	57	11.5	89	3	AAW33998 Ehrlichia
19	57	11.5	89	5	AAU73195 Human gra
20	56	11.3	86	5	AAU97011 Human GTP
21	55.5	11.2	68	6	ABJ19123 Pathogen
22	55	11.1	75	5	ABR07109 Human ORF
23	55	11.1	77	6	ADB06992 Alloiococ
24	54	10.9	80	4	AAW59349 Propionib
25	54	10.9	80	6	AAW55868 Propionib

25	52.5	10.6	33	5	ABP26511 Streptoco
27	52.5	10.6	56	6	AAW51869 Propionib
28	52.5	10.6	56	4	ABW48388 Propionib
29	52.5	10.6	70	4	ABG03427 Novel hum
30	52	10.5	80	4	ABG03792 Novel hum
31	52	10.5	95	4	ABG14805 Novel hum
32	52	10.5	95	4	ABG03922 Novel hum
33	51.5	10.4	41	5	ABJ05481 Human bre
34	51.5	10.4	78	7	ADD44488 Polypepti
35	51	10.3	47	6	ABO13991 Novel hum
36	51	10.3	48	2	AAW27586 Human sec
37	51	10.3	50	2	AAW38844 Human AX9
38	51	10.3	58	4	AAW55575 Propionib
39	51	10.3	58	6	AAW52094 Propionib
40	51	10.3	59	4	AAW84669 Human imm
41	51	10.3	65	5	ABP02952 Human ORF
42	51	10.3	77	4	AAW85096 Human imm
43	50.5	10.2	53	4	AAW21478 Peptide #
44	50.5	10.2	53	5	ABG4567 Human pep
45	50.5	10.2	80	4	AAW89190 Human imm

## ALIGNMENTS

RESULT 1  
AAW15359  
ID AAW15359 standard; protein; 76 AA.  
XX  
AC AAW15359;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #1793 encoded by probe for measuring cervical gene expression.  
XX  
KM Probe; human; microarray; gene expression; cervical epithelial cell;  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000670.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236358P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI, 2001-488901/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX  
PT gene expression in human cervical epithelial cells.  
XX  
PS Claim 27; SEQ ID NO 20185; 487bp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
XX  
CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded  
XX  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
XX  
CC can be used to produce a single exon microarray, which can be used for  
XX  
CC measuring human gene expression in a sample derived from human cervical  
XX  
CC epithelial cells. By measuring gene expression, the probes are therefore  
XX  
CC useful in grading and/or staging of diseases of the cervix, notably  
XX  
CC cervical cancer. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 76 AA;

Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 60  
DB 9 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 68

QY 61 LSQKLRQE 68  
DB 69 LSQKLRQE 76

RESULT 2  
ABB34356  
ID ABB34356 standard; peptide; 76 AA.

AC ABB34356;  
DT 04-FEB-2002 (first entry)

XX Peptide #1862 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.  
XX WO200157277-A2.

XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.  
XX Claim 27; SEQ ID NO 26991; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human fetal liver. The  
XX present sequence is a peptide encoded by a single exon nucleic acid probe  
XX of the invention. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 76 AA;

Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 60

DB 9 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 68  
QY 61 LSQKLRQE 68  
DB 69 LSQKLRQE 76

RESULT 3  
AAM27837  
ID AAM27837 standard; protein; 76 AA.

AC AAM27837;  
DT 17-OCT-2001 (first entry)

XX Peptide #1874 encoded by probe for measuring placental gene expression.  
XX Probe: microarray; human; placenta; antenatal diagnosis;  
XX Genetic disorder.

OS Homo sapiens.  
XX WO200157272-A2.

XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.  
XX Claim 27; SEQ ID NO 28106; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;  
XX see A191315-A157546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders

XX Sequence 76 AA;

Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 60  
DB 9 LEDSSILEVDGPAKKNVEFKAQKRGKGLCTVHREADPFWSLCTADMSLLEQSHSSPSLYIQC 68

QY 61 LSQKLRQE 68  
DB 69 LSQKLRQE 76

RESULT 4  
ABB29199  
ID ABB29199 standard; peptide; 76 AA.

AC ABB29199;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Peptide #1850 encoded by breast cell single exon nucleic acid probe.  
 XX  
 KW Human; microarray; single exon probe; gene expression; breast; disease;  
 cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 XX  
 PR 26-MAY-2000; 2000US-0207456P.  
 XX  
 PR 30-JUN-2000; 2000US-00608408.  
 XX  
 PR 03-AUG-2000; 2000US-0063236P.  
 XX  
 PR 21-SEP-2000; 2000US-0234687P.  
 XX  
 PR 27-SEP-2000; 2000US-0236359P.  
 XX  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WFI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes, useful  
 XX  
 PT for measuring gene expression in sample derived from human breast,  
 XX  
 PT comprises number of single exon nucleic acid probes.  
 XX  
 PS Claim 27; SEQ ID NO 12167; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BR 474 cells. The method involves contacting the  
 CC probes with a collection of detectably labelled nucleic acids derived  
 CC from mRNA of human breast, and then measuring the label bound to each  
 CC probe of the microarray. The probes are useful for verifying the  
 CC expression of the regions of genomic DNA predicted to encode proteins. They  
 CC are useful for gene discovery, and for determining predisposition and/or  
 CC prognosing breast disease. Gene expression analysis is useful for  
 CC assessing the toxicity of chemical agents on cells. The microarray of  
 CC this invention presents a far greater diversity of probes for measuring  
 CC gene expression, with far less bias than expressed sequence tag  
 CC microarrays. The method is suitable for rapid production of functional  
 CC information from genomic sequence. The present sequence is a peptide  
 CC encoded by a single exon nucleic acid probe of the invention. Note: The  
 CC sequence data for this patent did not form part of the invention.  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 76 AA:  
 XX  
 Query Match 71.0%; Score 353; DB 4; Length 76;  
 Best local Similarity 100.0%; Pred. No. 4,8e-36;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LENSILLETVDGAMNVEFFKQKRGKLCVHNEAFPMFLCTADMSLLEDSHSPSLYLQC 60  
 DB 9 LLESSLILEVDGAMNVEFFKQKRGKLCVHNEAFPMFLCTADMSLLEDSHSPSLYLQC 68  
 QY 61 LSGKLRQE 68  
 DB 69 LSGKLRQE 76  
 RESULT 5

AB91775  
ID ABB19775 standard; protein; 76 AA.  
XX  
AC ABB19775;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #1774 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KM cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN MO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00568408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR MPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts.  
XX  
CI Claim 15; SEQ ID NO 21545; 530bp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA215153-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. CC sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences

Sequence 76 AA:  
SQ

Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 LEDSSILEVDGPMKNVFFKAKRGGLCTVREADDFEWSLCTADMSLLBQSHSPSLIYQC 60  
DB 9 LEDSSILEVDGPAMKNVFKAQRKGLCTVREADDFWSLCTADMSSLBQSHSPSLIYLC 68  
QY 61 LSKXLRQE 68  
DB 69 LSKXLRF 76

RESULT 6  
AA67546  
ID AA67546 standard; protein; 76 AA.  
XX  
NC AA67546;  
XX

DT	XX	06-NOV-2001	(first entry)	
XX	XX			
XX	XX	Human bone marrow expressed probe encoded protein SEQ ID NO: 27852.		
XX	XX	Human, bone marrow expressed exon; gene expression analysis; probe;		
KW	KW	microarray; cancer; leukemia; lymphoma; myeloma.		
XX	XX			
OS	XX	Homo sapiens.		
XX	XX			
XX	XX	W0200157276-A2.		
XX	XX			
PD	XX	09-AUG-2001.		
XX	XX			
PF	XX	30-JAN-2001; 2001MO-US000668.		
XX	XX			
XX	XX	04-FEB-2000; 2000US-0180312P.		
FR	XX	26-MAY-2000; 2000US-0207456P.		
FR	XX	30-JUN-2000; 2000US-00608408.		
PR	XX	03-AUG-2000; 2000US-00633366.		
PR	XX	21-SEP-2000; 2000US-0234687P.		
PR	XX	27-SEP-2000; 2000US-0236559P.		
PR	XX	04-OCT-2000; 2000GB-00024266.		
XX	XX			
PA	XX	(MOLE-) MOLECULAR DYNAMICS INC.		
PI	XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	XX	WPI; 2001-488900/53.		
DR	XX			
XX	XX	Human genome-derived single exon nucleic acid probes useful for analyzing		
PT	XX	gene expression in human bone marrow.		
XX	XX			
PS	XX	Example 4; SEQ ID NO 27852; 658bp + sequence listing; English.		
XX	XX			
CC	XX	The present invention provides a number of single exon nucleic acid		
CC	XX	probes which are derived from genomic sequences expressed in the human		
CC	XX	bone marrow. They can be used to measure gene expression in bone marrow		
CC	XX	samples, which may enable the improved diagnosis and treatment of cancers		
CC	XX	such as lymphoma, leukemia and myeloma. The present sequence is a		
CC	XX	protein encoded by one of the probes of the invention		
XX	XX			
XX	XX	Sequence 76 AA:		
XX	XX			
XX	XX	Query Match	71.0%;	Score 353; DB 4; Length 76;
XX	XX	Best Local Similarity	100.0%;	Pred. No. 4.8e-36;
XX	XX	Matches 68; Conservative	0;	Mismatches 0; Indels 0; Gaps 0
QY	XX	1 LEQSSILEVDGPMKVKVERKAKRGGLCTVYHRAQDFPMSICTADMSILEQSHSPSLYLQC	60	
DB	XX	9 LEQSSILEVDGPMKVKVERKAKRGGLCTVYHRAQDFPMSICTADMSILEQSHSPSLYLQC	68	
QY	XX	61 LSQKTRQE	68	
DB	XX	69 LSQKTRQE	76	
XX	XX			
XX	XX	AA055153;		
XX	XX	AA055153		
XX	XX	AA055153 standard; protein; 76 AA.		
XX	XX			
DT	XX	05-NOV-2001	(first entry)	
XX	XX			
DE	XX	Human brain expressed single exon probe encoded protein SEQ ID NO: 27258.		
XX	XX	Human; brain expressed exon; gene expression analysis; probe; microarray;		
KW	XX	Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.		
XX	XX			
OS	XX	Homo sapiens.		
XX	XX			
XX	XX	W0200157275-A2.		
XX	XX			

```

FD 09-AUG-2001.
XX
EF 30-JAN-2001; 2001WO-US000667.
XX
FR 04-FEB-2000; 2000US-0180312P.
FR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
FR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
FR 27-SEP-2000; 2000US-0236359P.
FR 04-OCT-2000; 2000GS-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI, 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 27258; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
CC
SQ Sequence 76 AA;
XX
Query Match 71.0%; Score 353; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 4,8e-36;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 LEDSSILEVDGPMKNVVEFKQKRGICTVHREADFFNSLCTADMILLESHSPSLYLQC 60
DB 9 LEBSLSILEVDGPAMNVEFKQKRGICTVHREADFFWSLCTADMILLESHSPSLYLQC 68
OY 61 LSQKLKQE 68
DB 69 LSQKLKQE 76
XX
RESULT 8
ABG49189
ID ABG49189 standard; peptide; 76 AA.
XX
AC ABG49189;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 27837.
XX
KM Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KM hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
WO200157273-A2.
XX
PN 09-AUG-2001.
PD
PF 30-JAN-2001; 2001WO-US000664.
PP
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GS-00024263.
XX

```

PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488898/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.  
XX  
PS Claim 27; SEQ ID NO 27837; 658pp; English.  
XX  
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 1109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridizes at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 76 AA;  
XX  
Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 LEDSSILEVDGPAKNVFEKAKRGKGLCTVHRADFFWMSLCTADMSLLEQSHSPSYLQC 60  
DB 9 LEDSSILEVDGPAKNVFEKAKRGKGLCTVHRADFFWMSLCTADMSLLEQSHSPSYLQC 68  
XX  
QY 61 LSQKLRQE 68  
DB 69 LSQKLRQE 76  
XX  
RESULT 9  
AAM03118  
ID AAM03118 standard; protein; 76 AA.  
XX  
AC AAM03118;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Peptide #1800 encoded by probe for measuring breast gene expression.  
XX  
KW Probe; human; breast disease; breast cancer; development disorder;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US000661.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-476286/51.  
DR  
XX Novel single exon nucleic acid probe used to measuring gene expression in  
PT a human breast.  
PT  
XX  
PS Claim 27; SEQ ID NO 11858; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AAM00010-AAM10067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridizes at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 76 AA;  
XX  
Query Match 71.0%; Score 353; DB 4; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 LEDSSILEVDGPAKNVFEKAKRGKGLCTVHRADFFWMSLCTADMSLLEQSHSPSYLQC 60  
DB 9 LEDSSILEVDGPAKNVFEKAKRGKGLCTVHRADFFWMSLCTADMSLLEQSHSPSYLQC 68  
XX  
QY 61 LSQKLRQE 68  
DB 69 LSQKLRQE 76  
XX  
RESULT 10  
ABG37140  
ID ABG37140 standard; peptide; 76 AA.  
XX  
AC ABG37140;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human peptide encoded by genome-derived single exon probe SEQ ID 26805.  
XX  
KW Human; single exon probe; asthma; lung cancer; COPD; ILD; disease;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Kerns-Ry-Pudlak syndrome; sarcoidosis; pulmonary haemoiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagazer syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
XX

PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
PS  
XX Claim 27; SEQ ID NO 26805; 634bp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Kartagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 76 AA;  
Query Match 71.0%; Score 353; DB 5; Length 76;  
Best Local Similarity 100.0%; Pred. No. 4.8e-36;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LEQSSILEVDGPMKRVKFEKAKRGKCTVHEADPFWSLCTADWSLLEQSSISPSLYTQC 60  
DB 9 LEQSSILEVDGPMKRVKFEKAKRGKCTVHEADPFWSLCTADWSLLEQSSISPSLYTQC 68  
QY 61 LSOCKTQRE 68  
DB 69 LSOCKTQRE 76

XX  
XX 25-JAN-2001 (first entry)  
XX  
XX Eucalyptus grandis transcription factor protein sequence #407.  
XX  
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
XX homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
XX type 2 CysHis2; CCAAT box element; MYB.  
XX  
XX Eucalyptus grandis.  
XX  
XX MO200053724-A2.  
XX  
XX 14-SEP-2000.  
XX  
XX 09-MAR-2000; 2000MO-US006112.  
XX  
XX 11-MAR-1999; 99US-00266513.  
XX  
XX 18-AUG-1999; 99US-0149485P.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX Wood M, McGrath A, Shank WA, Glenn M;  
XX WPI; 2000-579369/54.  
XX  
XX  
XX Claim 8; Page 694; 747bp; English.  
XX  
XX The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
CC CysHis2, CCAAT box elements and MYB  
XX  
SQ Sequence 92 AA;  
Query Match 12.1%; Score 60; DB 3; Length 92;  
Best Local Similarity 28.4%; Pred. No. 23;  
Matches 19; Conservative 17; Mismatches 27; Indels 4; Gaps 2;  
QY 2 EDSSILEVDGPMKRVKFEKAKRGKCTVHEADPFWSLCTADWSLLEQSSISPSLYTQC 61  
DB 4 EKIKIKKIDNTLAVQVTSKRRRLI--KKAELSYLCAADVSLVFS-ATGKLYDFSS 59  
QY 62 SOKTQRE 68  
DB 60 SKRMKGE 66

RESULT 12  
AAG60552  
ID AAG60552 standard; protein; 71 AA.  
XX  
XX AAG60552;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 78442.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX

hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0128282P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 23-APR-1999; 99US-0130891P.  
XX 28-APR-1999; 99US-0131442P.  
XX 30-APR-1999; 99US-0132048P.  
XX 30-APR-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
XX 18-JUN-1999; 99US-0139750P.  
XX 18-JUN-1999; 99US-0139763P.  
XX 21-JUN-1999; 99US-0139817P.  
XX 22-JUN-1999; 99US-0139899P.  
XX 23-JUN-1999; 99US-0140353P.  
XX 23-JUN-1999; 99US-0140354P.  
XX  
XX 24-JUN-1999; 99US-0140695P.  
XX 28-JUN-1999; 99US-0140823P.  
XX 29-JUN-1999; 99US-0140991P.  
XX 30-JUN-1999; 99US-0141287P.  
XX 01-JUL-1999; 99US-0141842P.  
XX 01-JUL-1999; 99US-0142054P.  
XX 02-JUL-1999; 99US-0142055P.  
XX 06-JUL-1999; 99US-0142350P.  
XX 08-JUL-1999; 99US-0142803P.  
XX 09-JUL-1999; 99US-0142920P.  
XX 12-JUL-1999; 99US-0142977P.  
XX 13-JUL-1999; 99US-0143542P.  
XX 14-JUL-1999; 99US-0143624P.  
XX 15-JUL-1999; 99US-0144005P.  
XX 15-JUL-1999; 99US-0144085P.  
XX 16-JUL-1999; 99US-0144086P.  
XX 16-JUL-1999; 99US-0144325P.  
XX 19-JUL-1999; 99US-0144331P.  
XX 19-JUL-1999; 99US-0144332P.  
XX 19-JUL-1999; 99US-0144333P.  
XX 19-JUL-1999; 99US-0144334P.  
XX 19-JUL-1999; 99US-0144335P.  
XX 19-JUL-1999; 99US-0144352P.  
XX 20-JUL-1999; 99US-0144632P.  
XX 20-JUL-1999; 99US-0144884P.  
XX 21-JUL-1999; 99US-0144814P.  
XX 21-JUL-1999; 99US-0145066P.  
XX 21-JUL-1999; 99US-0145088P.  
XX 22-JUL-1999; 99US-0145089P.  
XX 22-JUL-1999; 99US-0145087P.  
XX 22-JUL-1999; 99US-0145089P.  
XX 22-JUL-1999; 99US-0145192P.  
XX 23-JUL-1999; 99US-0145145P.  
XX 23-JUL-1999; 99US-0145218P.  
XX 23-JUL-1999; 99US-0145224P.  
XX 26-JUL-1999; 99US-0145276P.  
XX 27-JUL-1999; 99US-0145913P.  
XX 27-JUL-1999; 99US-0145918P.  
XX 27-JUL-1999; 99US-0145919P.  
XX 28-JUL-1999; 99US-0145951P.  
XX 02-AUG-1999; 99US-0146386P.  
XX 02-AUG-1999; 99US-0146388P.  
XX 02-AUG-1999; 99US-0146389P.  
XX 03-AUG-1999; 99US-0147038P.  
XX 04-AUG-1999; 99US-0147204P.  
XX 05-AUG-1999; 99US-0147302P.  
XX 05-AUG-1999; 99US-0147192P.  
XX 06-AUG-1999; 99US-0147260P.  
XX 06-AUG-1999; 99US-0147303P.  
XX 06-AUG-1999; 99US-0147416P.  
XX 09-AUG-1999; 99US-0147935P.  
XX 09-AUG-1999; 99US-0148171P.  
XX 10-AUG-1999; 99US-0148319P.  
XX 11-AUG-1999; 99US-0148341P.  
XX 12-AUG-1999; 99US-0148565P.  
XX 13-AUG-1999; 99US-0148684P.  
XX 16-AUG-1999; 99US-0149175P.  
XX 17-AUG-1999; 99US-0149368P.  
XX 18-AUG-1999; 99US-0149422P.  
XX 20-AUG-1999; 99US-0149722P.  
XX 20-AUG-1999; 99US-0149723P.  
XX 20-AUG-1999; 99US-0149928P.  
XX 23-AUG-1999; 99US-0149902P.  
XX 23-AUG-1999; 99US-0149930P.  
XX 25-AUG-1999; 99US-0150566P.  
XX 26-AUG-1999; 99US-0150884P.  
XX 27-AUG-1999; 99US-0151065P.  
XX 27-AUG-1999; 99US-0151066P.  
XX 27-AUG-1999; 99US-0151080P.  
XX 30-AUG-1999; 99US-0151303P.  
XX 31-AUG-1999; 99US-0151338P.  
XX 01-SEP-1999; 99US-0151930P.

```
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155466P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158082P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159325P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160880P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 26-OCT-1999; 99US-0161820P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.7%; Score 58; DB 3; Length 71;
Best Local Similarity 45.2%; Pred. No. 23;
Matches 14; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

OY 50 SHSSPLVLOCUS----CKLRBERGTHPESG 76
Db 4 SNSSSLFLSCLSHTHKIRDRRGVRSKSTG 34

RESULT 13
AAG56185
ID AAG56185 standard; protein; 72 AA.
XX
AC AAG56185;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 72177.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
```

```
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131443P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132861P.
PR 11-MAY-1999; 99US-0132862P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134702P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 26-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138647P.
PR 14-JUN-1999; 99US-0139113P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140685P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
```



```
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144336P.
PR 20-JUL-1999; 99US-0144637P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147034P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 15-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 22-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151033P.
PR 30-AUG-1999; 99US-0151388P.
PR 31-AUG-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.

PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159284P.
PR 13-OCT-1999; 99US-0159285P.
PR 13-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.7%; Score 58; DB 3; Length 72;
Best Local Similarity 45.2%; Pred. No. 30;
Matches 14; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 50 SHSPSLYLQCLTS---CKLQKRGITIPGSG 76
DB 5 SNSSSLFLSCSSHTIRKIRDRGVKSTG 35

RESULT 14
ID AAM90887 standard; protein; 64 AA.
XX AAM90887;
AC AAM90887;
XX 07-NOV-2001 (first entry)
DT
DE Human immune/haematopoietic antigen SEQ ID NO:18460.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
XX WO200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
```

Thu Sep 9 07:46:00 2004

us-10-713-208-6\_copy\_253\_348.rag

Page 10

PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0232081P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234747P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 25-SEP-2000; 2000US-0234989P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246744P.  
PR 08-NOV-2000; 2000US-0246755P.  
PR 08-NOV-2000; 2000US-0246776P.  
PR 08-NOV-2000; 2000US-0246777P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0251989P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM,  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK63668.  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
PS Claim 11; SEQ ID NO 18480; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)

CC amino acid sequences given in AAM62170 to AAM91921. (1) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (1)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (1) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (1) by expressing inactive proteins or to  
CC supplement the patient's own production of (1). Additionally, (1)  
CC polynucleotides may be used to produce the secreted (1), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (1) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX Sequence 64 AA;  
SQ  
Query Match 11.6%; Score 57.5; DB 4; Length 64;  
Best Local Similarity 35.2%; Pred. No. 30;  
Matches 19; Conservative 7; Mismatches 23; Indels 5; Gaps 3;  
QY 35 FFW---SLCTADMSLLFQSHSPSLYLQCLSQKLFQEGTTPGSG-TRESKDM 84  
DB 4 FSWDRASLCHRGMSAMARWRLTATNSQLL-RLIKQENGVSFGSGCXELRSRH 56  
RESULT 15  
AAG61267  
ID AAG61267 standard; protein; 78 AA.  
XX  
XX AAG61267;  
AC  
XX  
XX 18-OCT-2000 (first entry)  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 79449.  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138647P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143432P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 18-JUL-1999; 99US-0144333P.  
PR 18-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.

```
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147431P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148655P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150866P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157177P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159283P.
PR 13-OCT-1999; 99US-0159284P.
PR 13-OCT-1999; 99US-0159285P.
PR 14-OCT-1999; 99US-0159349P.
PR 14-OCT-1999; 99US-0159349P.
PR 14-OCT-1999; 99US-0159349P.
PR 14-OCT-1999; 99US-0159349P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 16-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.

PR 22-OCT-1999; 99US-0160989P.
PR 23-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 11.6%; Score 57.5; DB 3; Length 78;
Best Local Similarity 32.2%; Pred. No. 38;
Matches 19; Conservative 6; Mismatches 19; Indels 15; Gaps 2;

QY 50 SNSSPELYQCLT-----OKLQERG-----TTPSGITESKDMFSSJGCIIL 93
Db 5 SNSSSLFLTCLSHYLITIKIRDRRGVSKSTGRATLGPFAVNDVTMFGGLVRLMM 63

Search completed: September 9, 2004, 06:15:00
Job time : 62 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:13:15 ; Search time 51 Seconds  
(without alignments)  
603.651 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_253\_348

Perfect score: 497  
Sequence: 1 LEDSSILEVDGPMKNVFEK.....ITTSKDNHFSGLGILDLVL 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 544098

Minimum DB seq length: 0  
Maximum DB seq length: 96

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/ECT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/ECTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	71.0	76	9 US-09-864-761-35073	Sequence 35073, A
2	63	12.7	50	12 US-10-424-599-178808	Sequence 178808, A
3	60.5	12.2	86	16 US-10-437-963-140300	Sequence 140300, A
4	60	12.1	85	12 US-10-424-599-207479	Sequence 207479, A
5	58.5	11.8	67	16 US-10-437-963-138338	Sequence 138338, A
6	58.5	11.8	87	12 US-10-424-599-208754	Sequence 208754, A
7	58.5	11.8	94	12 US-10-424-599-193827	Sequence 193827, A
8	57.5	11.6	86	16 US-10-437-963-191500	Sequence 191500, A
9	57.5	11.5	89	9 US-09-159-469-26	Sequence 26, Appl
10	57	11.5	89	12 US-09-993-108-26	Sequence 26, Appl
11	57	11.5	96	12 US-10-424-599-206718	Sequence 206718, A
12	57	11.5	78	16 US-10-437-963-102975	Sequence 102975, A
13	55	10.9	75	16 US-10-437-963-104068	Sequence 104068, A
14	54	10.8	54	12 US-10-424-599-189111	Sequence 189111, A
15	53.5	10.8	54	12 US-10-424-599-189111	Sequence 189111, A

16	53.5	10.8	94	12 US-10-424-599-187873	Sequence 187873, A
17	53	10.7	51	12 US-10-424-599-194305	Sequence 194305, A
18	53	10.7	52	12 US-10-424-599-199876	Sequence 199876, A
19	53	10.7	54	12 US-10-424-599-267605	Sequence 267605, A
20	53	10.7	79	16 US-10-437-963-102725	Sequence 102725, A
21	53	10.7	84	12 US-10-424-599-268479	Sequence 268479, A
22	52	10.5	45	12 US-10-424-599-225162	Sequence 225162, A
23	52	10.5	56	12 US-10-424-599-154704	Sequence 154704, A
24	52	10.5	79	16 US-10-437-963-109948	Sequence 109948, A
25	52	10.5	83	12 US-10-424-599-219827	Sequence 219827, A
26	52	10.5	85	16 US-10-437-963-173719	Sequence 173719, A
27	52	10.5	91	12 US-10-425-114-57053	Sequence 57053, A
28	51.5	10.4	41	14 US-10-074-475-241	Sequence 241, Appl
29	51.5	10.4	62	12 US-10-424-599-176492	Sequence 176492, A
30	51.5	10.4	86	12 US-10-424-599-280476	Sequence 280476, A
31	51	10.3	47	10 US-09-974-879-158	Sequence 158, Appl
32	51	10.3	47	12 US-10-621-401-158	Sequence 158, Appl
33	51	10.3	48	10 US-09-305-736-158	Sequence 158, Appl
34	51	10.3	48	11 US-09-818-683-158	Sequence 158, Appl
35	51	10.3	65	16 US-10-767-701-54657	Sequence 54657, A
36	51	10.3	75	16 US-10-437-963-141261	Sequence 141261, A
37	51	10.3	81	12 US-10-424-599-159132	Sequence 159132, A
38	51	10.3	91	12 US-10-424-599-158966	Sequence 158966, A
39	51	10.3	91	16 US-10-437-963-128297	Sequence 128297, A
40	50.5	10.2	53	9 US-09-864-761-48287	Sequence 48287, A
41	50.5	10.2	53	12 US-10-424-599-212918	Sequence 212918, A
42	50.5	10.2	62	12 US-10-424-599-247772	Sequence 247772, A
43	50.5	10.2	78	16 US-10-437-963-199589	Sequence 199589, A
44	50.5	10.2	96	12 US-10-424-599-220455	Sequence 220455, A
45	50	10.1	47	9 US-09-989-903-52	Sequence 52, Appl

## ALIGNMENTS

RESULT 1  
US-09-864-761-35073  
Sequence 35073, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-X-1  
CURRENT FILING DATE: 2001-05-23  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30

Thu Sep 9 07:46:00 2004

us-10-713-208-6\_copy\_253\_348.rapb

```

PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/224,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35073
LENGTH: 76
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007283.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.9
OTHER INFORMATION: EST_HUMAN HIT: A1139524.1, EVALU2 2.00e-38
US-09-864-761-35073
Query Match
Best Local Similarity 100.0%; Score 353; DB 9; Length 76;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 LEDSSLEVDGPAMKVEFAQCRGLCTVHREADPFWSLCTADMSLLEQSHSPSLYQC 60
DB 9 LEBSLLEVDGPAMKVEFAQCRGLCTVHREADPFWSLCTADMSLLEQSHSPSLYQC 68
QY 61 LSOCKLROE 68
DB 69 LSOCKLROE 76

RESULT 2
US-10-424-599-178808
Sequence 178808, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 178808
LENGTH: 50
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_132480C.1.pep
US-10-424-599-178808
Query Match
Best Local Similarity 12.7%; Score 63; DB 12; Length 50;
Matches 16; Conservative 7; Mismatches 12; Indels 22; Gaps 2;
```

```

QY 35 FFWSLCTADMSLLEQSHSPSLYQCLOKLRGERTIPSGITESHXDMFSSLGCT 91
DB 2 FFW-----LQKXSTSVFYEQ-----RSGGLTERSDHFNKNHCN 36

RESULT 3
US-10-437-963-140300
Sequence 140300, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 140300
LENGTH: 86
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(86)
OTHER INFORMATION: unsure at all Xaa locations
OTHER INFORMATION:
US-10-437-963-140300
Query Match
Best Local Similarity 12.2%; Score 60.5; DB 16; Length 86;
Matches 19; Conservative 10; Mismatches 23; Indels 7; Gaps 3;
```

```

QY 25 GLCTVHREADPFWSLCTADMSLLEQSHSPSLYQCLOKLRGERTIPSGITESHXDM 83
DB 19 GPCLRXREDMXGMCVCA-ARLLEALPRPS-VLETISRRLRRS-----GDDLDQEDVD 70

RESULT 4
US-10-424-599-207479
Sequence 207479, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 207479
LENGTH: 85
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_29379C.1.pep
US-10-424-599-207479
Query Match
Best Local Similarity 12.1%; Score 60; DB 12; Length 85;
Matches 23; Conservative 8; Mismatches 24; Indels 18; Gaps 3;
```



RESULT 9  
US-09-159-469-26  
Sequence 26, Application US/09159469  
Patent No. US2002006453A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
THERAPY OF EHRlichia INFECTION  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SPED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/159,469  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/106,582  
FILING DATE: 29-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 89 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Ehrlichia  
US-09-159-469-26

Query Match 11.5%; Score 57; DB 9; Length 89;  
Best Local Similarity 24.6%; Pred. No. 50;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

QY 33 ADFWNSLCTADMSLLEQSHSPSLYLQCLSGKLRQERGTTPSGITSKMHHFSSLCITL 92  
DB 28 SSFFFLYSTDF-LCTRSHSCPDY-----VTDKPOVFYAAGCVY 67

QY 93 L 93  
DB 68 I 68

RESULT 10  
US-09-798-042-26  
Sequence 26, Application US/09798042  
Patent No. US20020068343A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
FILE REFERENCE: 210121.439C7  
CURRENT APPLICATION NUMBER: US/09/798,042  
CURRENT FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 108  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 26  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Ehrlichia sp.  
US-09-798-042-26

Query Match 11.5%; Score 57; DB 9; Length 89;  
Best Local Similarity 24.6%; Pred. No. 50;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

QY 33 ADFWNSLCTADMSLLEQSHSPSLYLQCLSGKLRQERGTTPSGITSKMHHFSSLCITL 92  
DB 28 SSFFFLYSTDF-LCTRSHSCPDY-----VTDKPOVFYAAGCVY 67

QY 93 L 93  
DB 68 I 68

RESULT 11  
US-09-953-108-26  
Sequence 26, Application US/09953108  
Patent No. US20020086984A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
AND TREATMENT OF EHRlichia INFECTION  
FILE REFERENCE: 210121.439C8  
CURRENT APPLICATION NUMBER: US/09/953,108  
CURRENT FILING DATE: 2001-09-10  
NUMBER OF SEQ ID NOS: 110  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 26  
LENGTH: 89  
TYPE: PRT  
ORGANISM: Ehrlichia sp.  
US-09-953-108-26

Query Match 11.5%; Score 57; DB 12; Length 89;  
Best Local Similarity 24.6%; Pred. No. 50;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

QY 33 ADFWNSLCTADMSLLEQSHSPSLYLQCLSGKLRQERGTTPSGITSKMHHFSSLCITL 92  
DB 28 SSFFFLYSTDF-LCTRSHSCPDY-----VTDKPOVFYAAGCVY 67

QY 93 L 93  
DB 68 I 68

RESULT 12  
US-10-424-599-206718  
Sequence 206718, Application US/10424599  
Patent No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599



```

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 206718
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_28694C.1.pep
US-10-424-599-206718

Query Match
Best Local Similarity 28.7%; Pred. No. 56; Length 96;
Matches 25; Conservative 8; Mismatches 30; Indels 24; Gaps 4;

QY 10 DGPAMKVEFKAQKRGCTVREADFWSLCTADMSLL-EGSHSPSLYLQCLSQKLRQE 68
DB 23 DGSITKQV-----MSKMAEMINSLCGDEKIMRERTH---LTMWCSRMALRQK 68
QY 63 RGTIPGSGTBSKDMHPSLGCILIDV 95
DB 69 DGMWVECGIT-----TGCPPLTDV 86

RESULT 13
US-10-437-963-102975
; Sequence 102975, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 102975
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_100449C.1.pep
US-10-437-963-102975

Query Match
Best Local Similarity 11.1%; Score 55; DB 16; Length 78;
Matches 18; Conservative 13; Mismatches 23; Indels 20; Gaps 3;

QY 6 LIEVDGPAMKVEFKAQKRGCTVREADFWSLCTADMSLLEQSHSPSLYLQCLSQKL 65
DB 1 MKMDSLSIKGI-----ANNYLIVERFRKPFMFWRKSKIRRNQHSRP-----TQKV 48
QY 66 RQ-----ERGT 71
DB 49 KQPIYVIGNDKRG 62

RESULT 14
US-10-437-963-104068
; Sequence 104068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
```

```

; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 104068
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101437C.1.pep
US-10-437-963-104068

Query Match
Best Local Similarity 31.5%; Score 54; DB 16; Length 75;
Matches 17; Conservative 4; Mismatches 27; Indels 6; Gaps 2;

QY 37 WSLCTADMSLL---EGSHSPSLYLQCLSQKLRQERGTIPG--SGITBSKDMH 84
DB 12 WSCNMKRWVLLPDAPDGHRYSFQILSCGSLNLSMEERTTDGIRWITPRHLH 65

RESULT 15
US-10-424-599-189111
; Sequence 189111, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 189111
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(54)
; FEATURE:
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_141782C.1.pep
US-10-424-599-189111

Query Match
Best Local Similarity 10.8%; Score 53.5; DB 12; Length 54;
Matches 13; Conservative 10; Mismatches 19; Indels 3; Gaps 1;

QY 2 EDSLSLEVDGPAMKVEFKAQKRGCTVREADFWSLCTADMSL 46
DB 4 EKIHKKIDNASARHVTLLKRRRLFKISEPSF---LCADADAL 45

Search completed: September 9, 2004, 06:16:10
Job time : 52 secs
```





Db 28 SSFFFLYSTDf-LCTRSHSCPDAY-----VTDAKQVFAAGCV 67  
QY 93 L 93  
Db 68 I 68

RESULT 2  
US-08-821-324-26

Sequence 26, Application US/08821324  
Patent No. 6231869

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSER: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/821,324

FILING DATE: 21-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.439

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 89 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Ehrlichia

US-08-821-324-26

Query Match 11.5%; Score 57; DB 3; Length 89;  
Best Local Similarity 24.6%; Pred. No. 4.9; 17; Indels 20; Gaps 2;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

QY 33 ADFWSLCTADMSLLEQSHSBSPLYLOCLSQKROERGTFPGSGITKSMHFFSLGCL 92  
Db 28 SSFFFLYSTDf-LCTRSHSCPDAY-----VTDAKQVFAAGCV 67

QY 93 L 93  
Db 68 I 68

RESULT 3

US-09-295-028-26

Sequence 26, Application US/09295028

Patent No. 6277381

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond L.

APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION  
FILE REFERENCE: 210121.439C4  
CURRENT APPLICATION NUMBER: US/09/295,028  
CURRENT FILING DATE: 1999-04-20  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: Fast-Seq for Windows Version 3.0  
SEQ ID NO 26  
LENGTH: 89  
TYPE: PRP  
ORGANISM: Ehrlichia sp.  
US-09-295-028-26

Query Match 11.5%; Score 57; DB 3; Length 89;  
Best Local Similarity 24.6%; Pred. No. 4.9; 17; Indels 20; Gaps 2;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

QY 33 ADFWSLCTADMSLLEQSHSBSPLYLOCLSQKROERGTFPGSGITKSMHFFSLGCL 92  
Db 28 SSFFFLYSTDf-LCTRSHSCPDAY-----VTDAKQVFAAGCV 67

QY 93 L 93  
Db 68 I 68

RESULT 4

US-09-106-582-26

Sequence 26, Application US/09106582

Patent No. 6306402

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESSER: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/106,582

FILING DATE: 29-JUN-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.439C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 89 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Ehrlichia

US-09-106-582-26

Query Match 11.5%; Score 57; DB 4; Length 89;  
Best Local Similarity 24.6%; Pred. No. 4.9; 17; Indels 20; Gaps 2;  
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2;

[illegible]

RESULT 5  
ITS-09-15

US-09-159-469-26  
Sequence 26, Application US/09159469  
Patent No. 6607728  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND  
TITLE OF INVENTION: THERAPY OF EHRLICHIA INFECTION  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/159,469  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/106,582  
FILING DATE: 29-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.439C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-682-6031  
INFORMATION FOR SEQ. ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 89 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Ehrlichia

```

Query Match 11.5% Score 57; DB 4; Length 89;
Best Local Similarity 24.6% Pred. No. 4.9;
Matches 15; Conservative 9; Mismatches 17; Indels 20; Gaps 2
Cy 33 AAPPFSSCTADMSLLFQSHSSPSLYLQCLSKLRQENGTTPSGCTRESKDMFRSSLCGLT 92
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
28 SSGFYSLSTID-LTFRSHSCPAY-----VIDARQVYFAGCY 67

```

QY	93 L 93
Db	68 I 68

RESULT 6  
US-09-693-542-26

US-09-693-542-26

```

Sequence 26, Application US/09693542
Patent No. 667356
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.438C6
CURRENT APPLICATION NUMBER: US/09/693,542
CURRENT FILING DATE: 2000-10-20
NUMER OF SEQ ID NOS: 89
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 36
LENGTH: 89
TYPE: FRT
ORGANISM: Ehrlichia sp.
US-09-693-542-26

```

Query Match	11.5%;	Score 57;	DB 4;	Length 89;
Best Local Similarity	24.6%;	Pred. No. 4.9;		
Matches 15; Conservative	9;	Mismatches 17;	Indels 20;	Gaps 2

Qy	93	L	93
Db	68	I	68

## RESULT 7

```

US-09-621-976-7393
? Sequence 7393, Application US/09621976
? Patent No. 6639063
? GENERAL INFORMATION:
? APPLICANT: Dumas Milne Edwards, J.B.
? APPLICANT: Tobert, S.
? APPLICANT: Giordano, J.Y
? TITLE OF INVENTION: ESTs and Encoded Human Proteins
? FILE REFERENCE: GENSET-054PR2
? CURRENT APPLICATION NUMBER: US/09/621,976
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: Patent.pm
? SEQ ID NO 7393
? LENGTH: 75
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: UNSURE
? LOCATION: 4
? OTHER INFORMATION: Xaa = Ala, Asp, Glu, Val
US-09-621-976-7393

```

Query Match 10.4%; Score 51.5; DB 4; Length 75;  
Best Local Similarity 25.4%; Pred. No. 21;  
Matches 15; Conservative 12; Mismatches 23; Indels 9; Gaps 1.

## RESULT 8

US-09-187-789-52  
Sequence 52, Application US/09187789  
Patent No. 6340740  
GENERAL INFORMATION:  
APPLICANT: Alnemri; Emad S.  
APPLICANT: Fernandez-Alnemri, Teresa

;/ TITLE OF INVENTION: CASPASE-14, AN APOTOTIC PROTEASE, NUCLEIC ACID ENCODING  
;/ TITLE OF INVENTION: AND METHODS OF USE  
;/ FILE REFERENCE: 480140.434C1  
;/ CURRENT APPLICATION NUMBER: US/09/187,789  
;/ CURRENT FILING DATE: 1998-11-06  
;/ NUMBER OF SEQ ID NOS: 78  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 52  
;/ LENGTH: 47  
;/ TYPE: PRT  
;/ ORGANISM: Mus musculus  
;/ US-09-187-789-52

Query Match 10.1%; Score 50; DB 4; Length 47;  
Best Local Similarity 33.3%; Pred. No. 18;  
Matches 12; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 32 EADFFWSLCTADMSLLEQSHSSPSLYLQCLSQRLQ 67  
DB 10 EADFLGMATVLMCVSYRDPVNGTWYIQSLQSLRE 45

RESULT 9  
;/ Sequence 47, Application US/09139600  
;/ Patent No. 6432628  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Alnemir, Emdad S.  
;/ APPLICANT: Fernandez-Alnemir, Teresa  
;/ TITLE OF INVENTION: CASPASE-14, AN APOTOTIC PROTEASE, NUCLEIC ACID ENCODING  
;/ TITLE OF INVENTION: AND METHOD OF USE  
;/ FILE REFERENCE: 480140.434  
;/ CURRENT APPLICATION NUMBER: US/09/139,600  
;/ CURRENT FILING DATE: 1998-08-25  
;/ NUMBER OF SEQ ID NOS: 65  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 47  
;/ LENGTH: 47  
;/ TYPE: PRT  
;/ ORGANISM: Mus musculus  
;/ US-09-139-600-47

Query Match 10.1%; Score 50; DB 4; Length 47;  
Best Local Similarity 33.3%; Pred. No. 18;  
Matches 12; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 32 EADFFWSLCTADMSLLEQSHSSPSLYLQCLSQRLQ 67  
DB 10 EADFLGMATVLMCVSYRDPVNGTWYIQSLQSLRE 45

RESULT 10  
;/ US-09-107-532A-4733  
;/ Sequence 4733, Application US/09107532A  
;/ Patent No. 6583275  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Lynn A Doucette-Stamm and David Bush  
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
;/ ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
;/ NUMBER OF SEQUENCES: 7310  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
;/ STREET: 100 Beaver Street  
;/ CITY: Waltham  
;/ STATE: Massachusetts  
;/ COUNTRY: USA  
;/ ZIP: 02354  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: CD-ROM ISO9660  
;/ COMPUTER: PC  
;/ OPERATING SYSTEM: <Unknown>  
;/ SOFTWARE: ASCII  
;/ CURRENT APPLICATION DATA:

;/ APPLICATION NUMBER: US/09/107,532A  
;/ FILING DATE: 30-Jun-1998  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 60/085,598  
;/ FILING DATE: 14 May 1998  
;/ APPLICATION NUMBER: 60/051571  
;/ FILING DATE: July 2, 1997  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Arinleilo, Pamela Deneke  
;/ REGISTRATION NUMBER: 40,489  
;/ REFERENCE/DOCKET NUMBER: GTC-012  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (781)893-5007  
;/ TELEFAX: (781)893-8277  
;/ INFORMATION FOR SEQ ID NO: 4733:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 96 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
;/ HYPOTHEICAL: YES  
;/ ORIGINAL SOURCE:  
;/ ORGANISM: Enterococcus faecium  
;/ FEATURE:  
;/ NAME/KEY: misc feature  
;/ LOCATION: (8) LOCATION 1...96  
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 4733:  
;/ US-09-107-532A-4733

Query Match 10.1%; Score 50; DB 4; Length 96;  
Best Local Similarity 23.2%; Pred. No. 47;  
Matches 13; Conservative 15; Mismatches 24; Indels 4; Gaps 1;

QY 14 MNQVEFKQKGLCTVARE---ADFFWSLCTADMSLLEQSHSSPSLYLQCLSQRL 65  
DB 2 VAHMEIKKKKRLALFENDEQOIGEMTWSAGPDMIDHTFVDPAYRGCKLAKEL 57

RESULT 11  
;/ US-09-543-681A-5822  
;/ Sequence 5822, Application US/09543681A  
;/ Patent No. 6605709  
;/ GENERAL INFORMATION:  
;/ APPLICANT: GARY BRETON  
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRIN  
;/ TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
;/ FILE REFERENCE: 2709.1002-001  
;/ CURRENT APPLICATION NUMBER: US/09/543,681A  
;/ CURRENT FILING DATE: 2000-04-05  
;/ PRIOR APPLICATION NUMBER: US 60/128,706  
;/ PRIOR FILING DATE: 1999-04-09  
;/ NUMBER OF SEQ ID NOS: 8344  
;/ SEQ ID NO 5822  
;/ LENGTH: 63  
;/ TYPE: PRT  
;/ ORGANISM: Proteus mirabilis  
;/ US-09-543-681A-5822

Query Match 10.0%; Score 49.5; DB 4; Length 63;  
Best Local Similarity 30.4%; Pred. No. 31;  
Matches 17; Conservative 9; Mismatches 19; Indels 11; Gaps 2;

QY 32 EADFFWSLCTADMSLLEQSHSSPSLYLQCLSQRLQERGTTPSGITGSKXHFSS 87  
DB 5 ESNFFVIAIEDLTL---SVSLPCSVNLISHKDR-----SLYTKRKDLIIIS 49

RESULT 12  
;/ US-08-905-223-481  
;/ Sequence 481, Application US/08905223  
;/ Patent No. 6222029  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Edwards, Jean-Baptiste D.

APPLICANT: Duclert, Aymeric  
APPLICANT: Lacroix, Bruno  
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
NUMBER OF SEQUENCES: 503  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Knobb, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Win95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,223  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-0176  
TELEFAX: (619) 235-8550  
INFORMATION FOR SEQ ID NO: 481:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
ORIGINAL SOURCE:  
ORGANISM: Homo Sapiens  
TISSUE TYPE: Brain  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 24..-1  
IDENTIFICATION METHOD: Von Heine matrix  
OTHER INFORMATION: score 4.2  
OTHER INFORMATION: seq SLPALSLRASP/RX  
US-08-905-223-481

Query Match 9.9%; Score 49; DB 3; Length 82;  
Best Local Similarity 28.1%; Pred. No. 52;  
Matches 16; Conservative 11; Mismatches 28; Indels 2; Gaps 2;

CY 12 PAKNVEFAQKKGCTVHREADPFWSLCTADMSLLEQSHSSPSLIYQCLSQKLRQE 68  
DB 14 PALA-LSLASPRXLCTATKQKSGQNL-EDMGSGEQKADPPATEKTLLEKYLE 68

RESULT 13  
US-08-737-560A-9  
Sequence 9, Application US/08737560A  
Patent No. 5928893  
GENERAL INFORMATION:  
APPLICANT: KANG, Chang-Yuul  
APPLICANT: KIM, Joong-Gon  
TITLE OF INVENTION: MONOCLONAL ANTIBODY SPECIFIC FOR HUMAN  
TITLE OF INVENTION: 4-1BB AND CELL LINE PRODUCING SAME  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KANG, Chang-Yuul  
STREET: Professor Apt. Ka-302, #244-2, Bongchun 7-dong,  
STREET: Kwanaek-gu  
CITY: Seoul  
STATE: Seoul  
COUNTRY: Republic of Korea  
ZIP: 151-057  
ADDRESSEE: KIM, Joong-Gon  
STREET: Hanyang Apt. 72-1213, Apgujung-dong, Kangnam-gu

CITY: Seoul  
STATE: Seoul  
COUNTRY: Republic of Korea  
ZIP: 135-110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/AT  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,560A  
FILING DATE: 13-NOV-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 95-8176  
FILING DATE: 08-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
OTHER INFORMATION: VK23.32/CL  
US-08-737-560A-9

Query Match 9.8%; Score 48.5; DB 2; Length 96;  
Best Local Similarity 29.5%; Pred. No. 75;  
Matches 13; Conservative 7; Mismatches 17; Indels 7; Gaps 1;

CY 40 CTADMSL-----LEQSHSSPSLIYQCLSQKLRQERGTTIPGSG 76  
DB 23 CRASQSIDYLHWYQKSHSPRLIKYASQSIGIPSPFGSG 66

RESULT 14  
US-09-052-089A-5  
Sequence 5, Application US/09052089A  
Patent No. 6346605  
GENERAL INFORMATION:  
APPLICANT: Lee, Soo Y.  
TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER  
FAMILY, AND USES THEREOF  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSER: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,089A  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1

Thu Sep 9 07:46:00 2004

us-10-713-208-6\_copy\_253\_348.rai

Page 6

```
TELECOMMUNICATION INFORMATION:
/
/ TELEPHONE: 201-487-5800
/
/ TELEFAX: 201-343-1684
/
/ INFORMATION FOR SEQ ID NO: 5:
/
/ SEQUENCE CHARACTERISTICS:
/
/   LENGTH: 51 amino acids
/   TYPE: amino acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/
/ MOLECULE TYPE: peptide
/
/ HYPOTHETICAL: NO
/
/ FRAGMENT TYPE: internal
/
/ ORIGINAL SOURCE:
/
/   ORGANISM: Homo sapiens
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
/
/ US-09-052-089A-5
```

```
Query Match          9.6%; Score 47.5; DB 4; Length 51;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 9; Mismatches 10; Indels 7; Gaps 3;

QY      24 RGLCTVHEADPFMSLCTADMSLLEQSHSPSYLQCLSQ 63
Db      1 RALCTIC--SDFFDH--SRDVAAADCGH--TFHLQCLIQ 33
```

```
RESULT 15
US-09-621-976-4271
/ Sequence 4271, Application US/09621976
/ Patent No. 6639063
/
/ GENERAL INFORMATION:
/
/ APPLICANT: Dumas Milne Edwards, J.B.
/
/ APPLICANT: Jobert, S.
/
/ APPLICANT: Giordano, J.Y.
/
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/
/ FILE REFERENCE: GENSET.054PR2
/
/ CURRENT APPLICATION NUMBER: US/09/621,976
/
/ CURRENT FILING DATE: 2000-07-21
/
/ NUMBER OF SEQ ID NOS: 19335
/
/ SOFTWARE: Patent.pm
/
/ SEQ ID NO 4271
/
/ LENGTH: 86
/
/ TYPE: PRT
/
/ ORGANISM: Homo sapiens
/
/ FEATURE:
/
/ NAME/KEY: UNSURE
/
/ LOCATION: 47
/
/ OTHER INFORMATION: Xaa = *, Glu
/
/ US-09-621-976-4271
```

```
Query Match          9.6%; Score 47.5; DB 4; Length 86;
Best Local Similarity 31.0%; Pred. No. 88;
Matches 18; Conservative 9; Mismatches 24; Indels 7; Gaps 3;

QY      12 PAKNVEFFKQKRGCTVHEADPFMSLCTADMSLLEQSHSPSYLQCLSQKUR 66
Db      27 PSEKH--PFRASP--CFPHTPSDLVVQCFQEVTLVDSPFLESGVSPKLPCHTSELR 80
```

Search completed: September 9, 2004, 06:13:32  
Job time : 20 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:15:17 ; Search time 16 Seconds

(without alignments)  
450.898 Million cell updates/sec

Title: US-10-713-208-6\_copy\_62\_136

Sequence: 1 NHRIDLTKTKQKXKQSVQ.....SEAFIPQSTPEERYKKSKRP 75

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 23401

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	14.8	73	1 DNBPD8	DNA-binding protei
2	53	14.0	65	2 AE2575	hypothetical prote
3	50.5	13.3	71	2 P97824	50S ribosomal prot
4	50	13.2	57	2 T43742	ribosomal protein
5	50	13.2	57	2 AE1258	30S ribosomal prot
6	50	13.2	57	2 A11620	30S ribosomal prot
7	49	12.9	57	2 D69701	ribosomal protein
8	49	12.9	58	2 P89338	30S ribosomal prot
9	48.5	12.8	64	2 P84858	hypothetical prote
10	47.5	12.5	44	2 C34327	toponin T, fast s
11	47.5	12.5	72	2 S74974	hypothetical prote
12	46.5	12.3	64	2 T30097	hypothetical prote
13	46	12.1	62	2 T03340	gene e27 protein -
14	45.5	12.0	71	2 A71671	ribosomal protein
15	45	11.9	38	1 HMGS	exendin-1 - Mexica
16	45	11.9	67	2 F97099	hypothetical prote
17	45	11.9	69	2 B97745	hypothetical prote
18	44.5	11.7	69	2 C96021	hypothetical prote
19	43.5	11.5	54	2 T04034	hypothetical prote
20	43	11.3	44	2 T07452	hypothetical prote
21	43	11.3	47	2 E70227	hypothetical prote
22	43	11.3	58	2 G95164	ribosomal protein
23	43	11.3	60	2 AC9981	hypothetical prote
24	43	11.3	65	2 ARI590	hypothetical prote
25	43	11.3	66	2 P98030	30S ribosomal prot
26	43	11.3	73	2 T38402	hypothetical prote
27	43	11.3	75	2 H90913	hypothetical prote
28	43	11.3	75	2 AE2659	conserved hypochet
29	43	11.3	75	2 C97441	hypothetical prote

30	42.5	11.2	59	2 T03335	gene e32 protein -
31	42.5	11.2	67	2 D97745	hypothetical prote
32	42.5	11.2	72	2 H86660	hypothetical prote
33	42	11.1	50	2 H70224	hypothetical prote
34	42	11.1	64	2 B69172	sensory transducti
35	42	11.1	68	2 G84484	hypothetical prote
36	42	11.1	73	2 T39159	GTP-binding regula
37	41.5	10.9	67	2 S61817	M-like protein em
38	41.5	10.9	74	2 AF0545	probable secreted
39	41	10.8	56	2 B95096	hypothetical prote
40	41	10.8	58	2 G81896	hypothetical prote
41	41	10.8	65	2 AG1649	chymotrypsin inhib
42	41	10.8	72	2 B29537	hypothetical prote
43	41	10.8	72	2 S76555	hypothetical prote
44	41	10.8	73	2 I47089	keratin type II -
45	41	10.8	73	2 B70068	hypothetical prote

## ALIGNMENTS

## RESULT 1

DNBPD8  
DNA-binding protein Ner - phage D108

C/Species: phage D108  
C/Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 22-Jun-1999

C/Accession: A24680; S07931

R/Tollas, P.P.; Dubow, W.S.  
EMBO J. 4, 3031-3037, 1985

A/Title: The cloning and characterization of the bacteriophage D108 regulatory DNA-bind

A/Reference number: A24680; MUID:86055744; PMID:2998774

A/Accession: A24680

A/Molecule type: DNA

A/Residues: 1-73 <TOL>

A/Residues: 1-73 <MTZ>

A/Cross-references: EMBL:X03847; NID:915439; PIDN:CAA27475.1; PID:915441

C/Genetics:

A/Gene: ner  
C/Suprafamily: phage D108 DNA-binding protein

C/Keywords: DNA binding

Query Match 14.8%, Score 56; DB 1; Length 73;  
Best Local Similarity 28.4%; Pred. No. 57;

Matches 19; Conservative 11; Mismatches 31; Indels 6; Gaps 2;

QY	3	HRIIDTKIKQYKQSVQAGSYRNVIQALIOKSLKDPNNFREPVKKSIOSEAFIPQ 62
DB	13	HRADIVAEIRKRNMSIAELGRS--NHLSSTLKNALDK---RYPKAEKTIADALGNTFQ 66
QY	63	SIPERY 69
DB	67	DIWPERY 73

## RESULT 2

AE2575  
Hypothetical protein asr9028 [imported] - Nostoc sp. (strain PCC 7120) plasmid PCC7120e

C/Species: Nostoc sp. PCC 7120  
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C/Accession: AE2575

R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuch

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, M.;

DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A/Reference number: AB1807; MUID:21595285; PMID:11759840





```

Query Match      12.1%; Score 46; DB 2; Length 62;
Best Local Similarity 27.6%; Pred. No. 4.9e+02;
Matches 16; Conservative 10; Mismatches 22; Indels 10; Gaps 2.

13 KKKQSGAGCTGYRVL-QPALIKSLKDPNNRPEPVKKSIOESAFLPQSPPEERY 69
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :

```

A:Accession: A01555  
A:Molecule type: protein  
A:References: 1-38 <PAR>  
R:Vandermeers, A.J., Gourlet, P.; Vandermeers-Piret, M.C.; Gauthier, A.; De Neef, P.; Rattiner, U. Biochem. 164, 321-327, 1987  
A:Title: Chemical, immunological and biological properties of peptides like vasoactive intestinal peptide (VIP) from *Heloderma* (suspectum).  
A:Reference number: A37584; MUID:87190398; PMID:3569266  
A:Contents: annotation  
A:Note: reanalysis of peptide components in the venoms of *Heloderma horridum* and *H. s. eximium*-2 is the major peptide from *H. suspectum* venom (very small amounts of exenatide may have been misidentified)  
C:Comment: Exending is a venom component that are thought to bind to receptors for vasopressin  
G:In secretion of amylase.  
G:Superfamily: glucagon  
C:Keywords: duplication; secretagogue; venom  
F:1-38/Product: exenatide-1 [heliospectin I] #status experimental <Hs1>  
F:1-37/Product: heliospectin II #status experimental <Hs2>

Query Match	11.9%	Score 45	DB 1	Length 38
Best Local Similarity	47.4%	Pred. NO. 3.6e+02		
Matches	9	Conservative	4	Mismatches 6
				Indels 0
				Gaps 0
QY	8	KTRIKRYKOSVQAGTSYR	26	
		: : : : : : :		

Db 16 KIALOKYESTIGSSTSPR 34

Search completed: September 9, 2004, 06:18:32  
Job time : 18 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:14:06 ; Search time 9 Seconds  
(without alignments)  
433,918 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_62\_136  
Perfect score: 379  
Sequence: 1 NIHRLDKTKIKYKQSVQG.....SEAFLEPSIEPKRYKSKP 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 8856

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	14.8	73	1	NER_BPDI0
2	53	14.0	58	1	R821_STRPY
3	50.5	13.3	71	1	R129_RICCN
4	50	13.2	57	1	R821_LISMO
5	49	12.9	56	1	R821_BACSU
6	45.5	12.0	58	1	R821_STAMM
7	45.5	12.0	53	1	BFP_VIBFI
8	45.5	12.0	71	1	RL29_RICPR
9	45	11.9	38	1	EXEL_HELISU
10	44.5	11.7	75	1	TALA_CARMA
11	44	11.6	67	1	ZSS_MARTI
12	43.5	11.5	60	1	Y69A_CLOTE
13	43	11.3	58	1	R821_STRPN
14	43	11.3	73	1	VAUE_SCHPO
15	42.5	11.2	72	1	CHH6_PENMP
16	42	11.1	72	1	GRGE_CAVI
17	42	11.1	73	1	GRGB_HUMAN
18	41	10.8	65	1	R821_LACPL
19	41	10.8	62	1	R821_BUCBP
20	41	10.8	72	1	IC13_HORVU
21	41	10.8	72	1	Y738_SYNY3
22	41	10.8	73	1	YMOO_BACSU
23	40.5	10.7	69	1	GRGU_BOVIN
24	40.5	10.7	69	1	GRGU_HUMAN
25	40.5	10.7	69	1	GRGU_HUMAN
26	40	10.6	73	1	Y179_UREPA
27	40	10.6	46	1	Y609_BORBU
28	40	10.6	58	1	YBBU_BACSU
29	40	10.6	58	1	R821_LACIA
30	40	10.6	72	1	Y02B_BPT4
31	39.5	10.4	69	1	YFXK_AZOCA
32	39.5	10.4	70	1	ATPS_MAIZE
33	39	10.3	67	1	GBGD_HUMAN

## ALIGNMENTS

34	39	10.3	74	1	ATP9_MAIZE
35	39	10.3	74	1	Y295_ARCFU
36	38.5	10.2	54	1	YRKS_BACSU
37	38.5	10.2	56	1	Y097_NPVAC
38	38.5	10.2	67	1	ATP8_MOUSE
39	38.5	10.2	67	1	RS21_AQUAS
40	38	10.0	35	1	EXE2_HELISU
41	38	10.0	62	1	YME2_ARCFU
42	38	10.0	65	1	RS21_CHLFE
43	38	10.0	67	1	GBGD_MOUSE
44	38	10.0	71	1	ST1B_XENLA
45	38	10.0	72	1	IF12_PALSO
					P00840 zea mays (m
					O29947 archaeoglob
					P54446 bacillus su
					P41657 autotroph
					P03930 mus muscula
					O67028 aquifex aeo
					P04204 helodema s
					O28022 archaeoglob
					O8K570 chlorobium
					O9JmT3 mus musculu
					O09005 xenopus lae
					O8Kx77 ralatonia s

RESULT 1	NER_BPDI0	STANDARD;	PRT;	73 AA.
ID	NER_BPDI0			
AC	P06803;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	DNA-binding protein Ner.			
GN	NER.			
OS	Bacteriophage D108.			
OC	Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;			
OC	Mu-like viruses.			
OK	NCBI_TaxID=10671;			
OK	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86055744; PubMed=2998774;			
RA	Tollas P.P., Dubow M.S.;			
RT	"The cloning and characterization of the bacteriophage D108			
RL	regulatory DNA-binding protein ner.";			
RL	EMBO J. 4:3031-3037(1985).			
RM	(2)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86232621; PubMed=3012481;			
RA	Mizunuchi M., Weisberg R.A., Mizunuchi K.;			
RT	"DNA sequence of the control region of phage D108: the N-terminal			
RT	amino acid sequences of repressor and transposase are similar both in			
RT	phage D108 and in its relative, phage Mu.";			
RL	Nucleic Acids Res. 14:3813-3825(1986).			
CC	- FUNCTION: BINDS TO THE REGION RIGHT AFTER THE START OF			
CC	TRANSCRIPTION OF THE PROMOTER REGION OF THE EARLY GENES.			
CC	- SIMILARITY: BELONGS TO THE NER FAMILY OF TRANSCRIPTIONAL			
CC	REGULATORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M26291; AAA32206.1; -			
DR	EMBL; X03847; CA27475.1; -			
DR	PIR; A24680; DNEPDS.			
DR	HSSP; P06020; INEQ.			
KW	Transcription regulation; DNA-binding.			
FT	DNA BIND 28 47 H-T-H MOTIF (PROBABLE).			
SQ	SEQUENCE 73 AA; 8534 MW; 898DFD38E1409025 CRC64;			
Query Match	14.8%; Score 56; DB 1; Length 73;			
Best Local Similarity	28.4%; Pred. No. 36;			
Matches	19; Conservative 11; Mismatches 31; Indels 6; Gaps 2;			
QY	3 HRIDLTKIKYKQSVQAGTSYRNVIQAATQSLKDPSPNFRFBPVKKSIGSEAFLLPQ 62			
DB	13 HRADIVAELEKRRNLSIAELGRS--NHLSSTLKNALDK---FYPAEKTIADALGNTQ 66			







RA Kobayashi Y.;  
 RT "Systematic sequencing of the 263 kb 210 degrees-232 degrees region of  
 RT the *Bacillus subtilis* genome containing the skin element and many  
 RT sporulation genes.";  
 RL Microbiology 142:3103-3111(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97175542; PubMed=9023197;  
 RA Homuth G., Masuda S., Mogk A., Kobayashi Y., Schumann W.;  
 RT "The dnaK operon of *Bacillus subtilis* is heptacistronic.";  
 RL J. Bacteriol. 179:1153-1164(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Aevedo V., Bertorello M.G., Bessières P., Bolotin A., Borcherdt S.,  
 RA Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Cornerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertlich K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gilm S.Y., Glaeser P., Gollighly E.V., Grandi G.,  
 RA Gillespie G., Guy B.J., Haga K., Haeleth U., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blandhard M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lamber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Médigue C.,  
 RA Medina N., Meliade R.P., Mizuno M., Moestl D., Nakai S., Noack M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Pressner E., Pulic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schoeter R., Scofield F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Solio B.,  
 RA Sokorin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takauchi V., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Toseko M., Uchiyama S., Vandembol M., Vannier F., Vassartoli A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Weissenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yatsunaga K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";  
 RT Nature 390:249-256(1997).  
 RN [4]  
 RP SEQUENCE OF 1-3;  
 RX MEDLINE=82219212; PubMed=6806564;  
 RA Higo K.-I., Otake E., Osawa S.;  
 RT "Purification and characterization of 30S ribosomal proteins from  
 RT *Bacillus subtilis*: correlation to *Bscherichia coli* 30S proteins.";  
 RL Mol. Genet. 185:239-244(1982).  
 CC -1- SIMILARITY: Belongs to the S21P family of ribosomal proteins.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D84433; BAA12470.1; -  
 DR EMBL: D87217; BAA12082.1; -  
 DR EMBL: Z99117; CAB14483.1; -  
 DR FTR; D69701; D69701.  
 DR Subtilin; BG1648; ipso.  
 DR HAMAP; MF\_00358; -1.  
 DR InterPro; IPR001911; Ribosomal\_S21.  
 DR Pfam; PF0165; Ribosomal\_S21; 1.  
 DR PRINTS; PR00976; RIBOSOMALS21.

DR ProDom; PD005521; Ribosomal\_S21; 1.  
 DR TIGRFAMs; TIGR00030; S21P; 1.  
 DR PROSITE; PS01181; RIBOSOMAL\_S21; 1.  
 KW Ribosomal protein; Complete proteome.  
 FT INIT\_MET 0  
 FT CONFLICT 0  
 SQ SEQUENCE 56 AA; 6699 MW; 221ED78138C8FBE4 CRC64;  
 S-> PRX (IN REF. 4).  
 Query Match 12.9%; Score 49; DB 1; Length 56;  
 Best Local Similarity 28.8%; Pred. No. 1,3e+02;  
 Matches 15; Conservative 12; Mismatches 13; Indels 12; Gaps 2;  
 QY 7 LKTKIQKYGQGVQAGTGYRNVLQALQKLDPSNNFREPVKKSIOESSEA 58  
 DB 11 LEDALRRKRSVSKTGT-----LQEARK--REFEYKPSVKRKKKSEA 50  
 RESULT 6  
 RS21\_STAMP STANDARD; PRT; 58 AA.  
 ID RS21\_STAMP  
 AC Q99T81;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 30S ribosomal protein S21.  
 GN RPSU OR SVA1575 OR SVA104 OR MW1527 OR SE1262.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
 OS Staphylococcus aureus (strain N315),  
 OS Staphylococcus aureus (strain MW2), and  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 NX NCB1\_TaxId158878, 158879, 196620, 1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus; STRAIN=MU50 / ATCC 700699, and N315;  
 RX MEDLINE=23111952; PubMed=11418146;  
 RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hoshiyama A.,  
 RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.I., Katto C.,  
 RA Sekimizu K., Hirakawa H., Kumara S., Goto S., Yabuzaki J.,  
 RA Kaneshita M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiratsuku K.;  
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.";  
 RT Lancet 357:1225-1240(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. aureus; STRAIN=MW2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiratsuku K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. epidermidis; STRAIN=ATCC 12228;  
 RX PubMed=12950922;  
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
 RA Qin Z.-Q., Mao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
 RT *Staphylococcus epidermidis* strain (ATCC 12228)." ;  
 RL Mol. Microbiol. 49:1577-1593(2003).  
 CC -1- SIMILARITY: Belongs to the S21P family of ribosomal proteins.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entitlee requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AP003362; BAB57737.1; -
DR  EMBL; AP003134; BAB42667.1; -
DR  EMBL; AP004827; BAB35392.1; -
DR  EMBL; AE016748; AAC04861.1; -
DR  FIR; F89938; F89938.
DR  HAMAP; MF_00358; -; 1.
DR  InterPro; IPR001911; Ribosomal_S21.
DR  Pfam; PF01165; Ribosomal_S21; 1.
DR  PRINTS; PR00976; RIBOSOMAL_S21.
DR  ProDom; PD005521; Ribosomal_S21; 1.
DR  TIGRFAMs; TIGR00030; S21P; 1.
DR  PROSITE; PS01181; RIBOSOMAL_S21; 1.
KW  Ribosomal protein; Complete proteome.
SQ  SEQUENCE 58 AA; 6972 MW; E3AE05B6AB7098AE CRC64;

Query Match
Best Local Similarity 12.9%; Score 49; DB 1; Length 58;
Matches 16; Conservative 11; Mismatches 13; Indels 12; Gaps 2;

Cc  7 LKTKYKQKQSVQAGTSYRNVLQAAIQKSLKPSNNFREBPVKKSIOSEEA 58
Cc  12 LEDALRRFRKRSVSKSGT-----IOEVRK---REFYKFSYKRRKXSEA 51

Db  12 LEDALRRFRKRSVSKSGT-----IOEVRK---REFYKFSYKRRKXSEA 51

RESULT 7
BFP_VIBFI
ID_BFP_VIBFI STANDARD; PRT; 53 AA.
AC P80893;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DR Blue fluorescence protein (BFP) (Fragment).
OS Vibrio fischeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=668;
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 33715 / Y-1;
RX MEDLINE=97326008; PubMed=9183020;
RA Fetushkov V.N., Lee J.;
RT "Purification and characterization of flavoproteins and cytochromes
RT from the yellow bioluminescence marine bacterium Vibrio fischeri
RT strain Y1.";
RL Eur. J. Biochem. 245:790-796 (1997).
CC -1- COFACTOR: 6,7-dimethyl-8-(1'-D-ribityl)lumazine, 6-methyl-7-oxo-
CC 8-(1'-D-ribityl)lumazine, or riboflavin; noncovalently bound.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: TO BIOLUMINESCENCE ANTENNA PROTEINS LUXL (LUMP) AND
CC LUXY (YFP) AND TO RIBOFLAVIN SYNTHASE ALPHA CHAIN.
DR InterPro; IPR001783; Lum binding.
DR PROSITE; PS00693; LUM_BINDING; PARTIAL.
KW Luminescence; Flavoprotein.
FT NON TER 53
SQ SEQUENCE 53 AA; 5658 MW; 744ED8F0F36789CC CRC64;

Query Match
Best Local Similarity 12.0%; Score 45.5; DB 1; Length 53;
Matches 15; Conservative 9; Mismatches 22; Indels 5; Gaps 2;

Cc  14 YKQSVQAGTSYRNVLQAAIQKSLKPSNNFREBPVKKSIOSEEAFLPQSI 64
Cc  2 FKGNVGVGT-VENIDKGAKFQSLRGVS---LPIPADLQSLDIFPEDI 47

Db  2 FKGNVGVGT-VENIDKGAKFQSLRGVS---LPIPADLQSLDIFPEDI 47

RESULT 8
RL29_RICPR STANDARD; PRT; 71 AA.
ID_RL29_RICPR
AC Q9ZCR3;

```

```

DR  30-MAY-2000 (Rel. 39, Created)
DR  30-MAY-2000 (Rel. 39, Last sequence update)
DR  28-FEB-2003 (Rel. 41, Last annotation update)
DE  50S ribosomal protein L29.
GN  RPMC OR Rp65l.
OS  Rickettsia prowazekii.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Rickettsiaceae; Rickettsiidae; Rickettsia.
OX  NCBI_TaxID=782;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Madrid E;
RX  MEDLINE=99039499; PubMed=9823893;
RA  Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA  Sickeritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA  Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT  "The genome sequence of Rickettsia prowazekii and the origin of
RT  mitochondria.";
RL  Nature 396:133-140 (1998).
CC  -1- SIMILARITY: Belongs to the L29P family of ribosomal proteins.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AJ235272; CAA15091.1; -
DR  FIR; A71671; A71671.
DR  HAMAP; MF_00374; -; 1.
DR  InterPro; IPR001854; Ribosomal_L29.
DR  Pfam; PF00831; Ribosomal_L29; 1.
DR  TIGRFAMs; TIGR00012; L29; 1.
DR  PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW  Ribosomal protein; Complete proteome.
SQ  SEQUENCE 71 AA; 8406 MW; D3D06C34DA617C3E CRC64;

Query Match
Best Local Similarity 12.0%; Score 45.5; DB 1; Length 71;
Matches 17; Conservative 10; Mismatches 33; Indels 7; Gaps 1;

Cc  10 KIQKTKQKQSVQAGTSYRN-----VLQAAIQKSLKPSNNFREBPVKKSIOSEEAFLPQ 62
Cc  5 KLRKSKSTETIEELYNLKLKKEFLNRFQALGELKNTSRSPVSKSIARIKTELTK 64

Db  5 KLRKSKSTETIEELYNLKLKKEFLNRFQALGELKNTSRSPVSKSIARIKTELTK 64

Cc  63 SIPHERY 69
Cc  65 RNSSEY 71

Db  65 RNSSEY 71

RESULT 9
EXEL_HELISU
ID_EXEL_HELISU STANDARD; PRT; 38 AA.
AC P04203;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE  Exendin-1 (Heliospectins I and II).
OS  Heloderma spectum (Gila monster).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Lepidodermata; Squamata; Scleroglossa; Anguilliformes; Helodermatidae;
OC  Heloderma.
OX  NCBI_TaxID=8554;
RN  [1]
RP  SEQUENCE.
RC  TISSUE=Venom;
RX  MEDLINE=85006896; PubMed=6207171;
RA  Parker D.S., Rautman J.-P., O'Donohue T.L., Bledsoe M., Yoshida H.,
RA  Pisano J.U.;
RT  "Amino acid sequences of heliospectins, new members of the glucagon
RT  superfamily, found in Gila monster venom.";

```

```

RL J. Biol. Chem. 259:11751-11755(1984).
CC -1- FUNCTION: Has a VIP/secretin-like biological activity.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: Belongs to the glucagon family.
DR InterPro: IPR000532; Glucagon.
DR Pfam: PF00123; hormone2; 1.
DR SMART: SM00070; GLUCA; 1.
DR PROSITE: PS00260; GLUCAGON; 1.
KW Glucagon family; Toxin.
FT VARIANT 38 54275ECFC36631A4 CRC64;
SQ SEQUENCE 38 AA; 4096 MW; 54275ECFC36631A4 CRC64;

Query Match 11.8%; Score 45; DB 1; Length 38;
Best Local Similarity 47.4%; Pred. No. 2.2e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 8 KTKIKKQKQVQAGTSYR 26
DB 16 KLVQKYLESLIGSISPR 34

RESULT 10
TALA CARMA STANDARD; PRT; 75 AA.
AC P80427;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transaldolase (EC 2.2.1.2) (Fragments).
OS Carcinus maenas (Common shore crab) (Green crab).
CC Eukaryote; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
CC Eudachyura; Portunidea; Portunidae; Carcinus.
CX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RA Lechaise F., Somme G., Carpentier G., Granjeon E., Webster S.,
RA Baghdassarian D.;
RT "A transaldolase. An enzyme implicated in crab steroidogenesis.";
RL Endocrine 5:23-32(1996).
CC -1- FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
CC METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY. MAY PLAY A ROLE IN
CC THE CONVERSION OF STEROIDS INTO ECYSTEROIDS VIA NADPH.
CC -1- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glucose 6-phosphate
CC -1- 3-phosphate = D-erythrose 4-phosphate + D-fructose 6-phosphate.
CC -1- PATHWAY: Pentose phosphate pathway; nonoxidative part.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN Y-ORGANS.
CC -1- PTM: Phosphorylated.
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN IS:
CC 6.8, ITS MW IS: 36.2 kDa.
CC -1- SIMILARITY: Belongs to the transaldolase family. Subfamily 1.
DR InterPro: IPR001585; Transaldolase.
DR PROSITE: PS01054; TRANSALDOLASE_1; 1.
DR PROSITE: PS00958; TRANSALDOLASE_2; PARTIAL.
KW Transferase; Pentose shunt; Phosphorylation.
FT NON TER 1 1
FT NON_CONS 17 18
FT NON_CONS 27 28
FT NON_CONS 34 35
FT NON_CONS 48 49
FT NON_CONS 64 65
FT NON TER 75 75
SQ SEQUENCE 75 AA; 8818 MW; BD844CEFAEB50559 CRC64;

Query Match 11.7%; Score 44.5; DB 1; Length 75;
Best Local Similarity 30.6%; Pred. No. 5.1e+02;
Matches 15; Conservative 11; Mismatches 18; Indels 5; Gaps 2;
```

```

RESULT 11
2S5_MATST STANDARD; PRT; 67 AA.
ID 2S5_MATST
AC P17718;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Matleuccin light and heavy chains (Fragments).
OS Matleuccia struthiopteris (Ostrich fern).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Nonillidomopses; Filicophyta; Filicopsida; Filicales; Dryopteridaceae;
OC Matleuccia.
CX NCBI_TaxID=3277;
RN [1]
RP SEQUENCE.
RA TISSUE=spore;
RX MEDLINE=90382429; PubMed=2401286;
RA Roedin J., Raek L.;
RT "Characterization of matleuccin, the 2.2S storage protein of the
RT ostrich fern. Evolutionary relationship to angiosperm seed storage
RT proteins.";
RL Bur. J. Biochem. 192:101-107(1990).
CC -1- FUNCTION: This is a 2S seed storage protein.
CC -1- SUBUNIT: HETERODIMER OF A SMALL AND A LARGE CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: PROTEIN BODIES.
CC -1- SIMILARITY: Belongs to the 2S seed storage albumins family.
DR PIR: S11350; S11350.
KW Seed storage protein.
FT CHAIN 1 >25 LIGHT CHAIN.
FT NON_CONS 25 26
FT CHAIN 26 >54 HEAVY CHAIN.
FT NON_CONS 54 55
FT CHAIN <55 >67 HEAVY CHAIN.
FT NON TER 67 67
SQ SEQUENCE 67 AA; 7961 MW; 141086B9B22AC55C CRC64;

Query Match 11.6%; Score 44; DB 1; Length 67;
Best Local Similarity 28.6%; Pred. No. 5e+02;
Matches 10; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 2 IHRIDLTKIKQKQVQAGTSYRNLQAIQKS 36
DB 12 LHCQDLRRDCARRSSENGGSERRKQQLRACDEDS 46

RESULT 12
Y99A_CLOTE STANDARD; PRT; 60 AA.
ID Y99A_CLOTE
AC P60083;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0291 protein CTC01690.1.
GN CTC01690.1.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Bruggemann H., Baumer S., Pricke W.F., Wierzer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias K., Meixl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the UPF0291 family.
```

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AE015941; -; NOT ANNOTATED; CDS.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 60 AA; 7133 MW; 3AB07AE4806E3505 CRC64;  
  
Query Match 11.5%; Score 43.5; DB 1; Length 60;  
Best Local Similarity 34.3%; Pred. No. 5e+02;  
Matches 12; Conservative 8; Mismatches 14; Indels 1; Gaps 1;  
  
14 YKSVQAGTGYRNVLOAI-QSKLDPNNFREP 47  
14 YKSKKEGTEBEKEEQALREKYLKRNFRNQ 48  
  
RESULT 13  
RS21 STRPN STANDARD; PRT; 58 AA.  
AC Q97017;  
DT 28-FEB-2003 (rel. 41, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE 30S ribosomal protein S21.  
GN RPSU OR SP1414 OR SPRI271.  
OS Streptococcus pneumoniae, and  
OS Streptococcus pneumoniae (strain ATCC BAA-295 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=1313, 171101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC BAA-334 / TIGR4;  
MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radtke D.,  
RA Holtzapple E., Khouri H., Wolf A.M., Utterberg T.R., Hansen D.,  
RA McDonald L.A., Feldblyum T.V., Angiulli S., Dickinson T., Hickey E.K.,  
RA Holt I.F., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae".  
RL Science 293:498-506 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC BAA-295 / R6.  
MEDLINE=2149245; PubMed=11544234;  
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Buxett S.,  
RA DeHoff B.S., Eklund S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Macsushina P.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.D., Lu J., Lagace R.E.,  
RA Mahren S.M., McNamey M., Mcleaster K., Mundy C.W., Nicks T.I.,  
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,  
RA Sun F.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,  
RA Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6".  
RL J. Bacteriol. 183:5709-5717 (2001).  
CC -1- SIMILARITY: Belongs to the S21P family of ribosomal proteins.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

-----  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AE007438; AAK75512.1; -  
DR EMBL; AE008438; AAL00075.1; ALT\_INIT.  
DR PIR; G95164; G95164.  
DR TIGR; SP1414; -  
DR HAMAP; MF\_00358; -; 1.  
DR InterPro; IPR001911; Ribosomal\_S21.  
DR Pfam; PF01165; Ribosomal\_S21; 1.  
DR PRINTS; PR00976; RIBOSOMALS21.  
DR ProDom; PD005521; Ribosomal\_S21; 1.  
DR TIGRFAMs; TIGR00030; S21P; 1.  
DR PROSITE; PS01181; RIBOSOMAL\_S21; 1.  
KW Ribosomal protein; Complete proteome  
SQ SEQUENCE 58 AA; 6998 MW; F7C36B5B86307C1 CRC64;  
  
Query Match 11.3%; Score 43; DB 1; Length 58;  
Best Local Similarity 27.5%; Pred. No. 5.4e+02;  
Matches 14; Conservative 12; Mismatches 13; Indels 12; Gaps 2;  
  
7 LKTKQKQSVQAGTGYRNVLOAIQSKLDPNNFREP 57  
12 LDDALRRFRRAVTRAGT-----LQETRK--REFYKPSYKRRKSE 50  
  
RESULT 14  
YABE SCHPO STANDARD; PRT; 73 AA.  
ID YABE SCHPO  
AC Q10167;  
DT 01-OCT-1996 (rel. 34, Created)  
DT 01-OCT-1996 (rel. 34, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE Hypothetical protein C26A3.14c in chromosome I.  
GN SPAC26A3.14C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Hornsby I., Howarth S., Huckle E.J., Hunt S., Jagers K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skellon J., Simmonds R., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Roben J., Grymptre B.,  
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Hojzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Mambutt R., Punelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Mottier S.,  
RA Calbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Inoue G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe".  
RL Nature 415:871-880 (2002).  
CC -1- SIMILARITY: Belongs to the S21P family of ribosomal proteins.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).

DR EMBL; Z69240; CAA93237.1; -  
DR PIR; T38402; T38402.  
DR GenBank; SPAC26A3.1ac; -  
KW Hypothetical protein.  
SQ SEQUENCE 73 AA; 8249 MW; B56C52B7B3435411 CRC64;

Query Match 11.3%; Score 43; DB 1; Length 73;  
Best Local Similarity 27.0%; Pred. No. 6.9e+02;  
Matches 10; Conservative 7; Mismatches 10; Indels 10; Gaps 1;

OY 2 IHRDLTKTKIOKY-----KQSVQAGACTSYRNV 28  
DB 37 INEDVKTAVKTLNFGEMAFDOSAFDLSSTWFKCV 73

## RESULT 15

CHH6\_PENJP STANDARD; PRT; 72 AA.  
AC P81770;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Crustacean hyperglycemic hormone 6 (Pej-SGP-VI).  
OS Pennaeus japonicus (Kuruma prawn).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Pennaeoidea;  
OC Pennaeidae; Marsupennaeus.  
OX NCBI\_TaxID=27405;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Sinus gland;  
RX MEDLINE=97353924; PubMed=9210164;  
RA Yang W.-T., Aida K., Nagasawa H.;  
RT "Amino acid sequences and activities of multiple hyperglycemic  
RL peptides from the kuruma prawn, Pennaeus japonicus.";  
CC -1- FUNCTION: Hormone found in the sinus gland of isopods and decapods  
CC which controls the blood sugar level. Has a secretagogue action  
CC over the amylase released from the midgut gland. May act as a  
CC stress hormone and may be involved in the control of molting and  
CC reproduction.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE MEDULLA TERMINALIS X-ORGAN IN  
CC THE EYESTALKS AND TRANSPORTED TO THE SINUS GLAND WHERE THEY ARE  
CC STORED AND RELEASED.  
CC -1- SIMILARITY: Belongs to the arthropod CHH/MIH/GIH/VIH hormone  
CC family.  
DR InterPro; IPR001166; CHH\_MIH\_GIH.  
DR Pfam; PF01147; Crust\_neurohorm; 1.  
DR PRINTS; PR00550; HYPERGLYCEMIC.  
DR PROSITE; PS01250; CHH\_MIH\_GIH; 1.  
KW Neuropeptide; Hormone; Glucose metabolism; Amidation.  
FT DISULFID 7 43 BY SIMILARITY.  
FT DISULFID 23 39 BY SIMILARITY.  
FT DISULFID 26 52 BY SIMILARITY.  
FT MOD RES 72 72 AMIDATION.  
SQ SEQUENCE 72 AA; 8311 MW; D1FC924A33545ED CRC64;

Query Match 11.2%; Score 42.5; DB 1; Length 72;

Best Local Similarity 29.6%; Pred. No. 7.7e+02;  
Matches 16; Conservative 4; Mismatches 23; Indels 11; Gaps 1;

OY 25 YRNVLTQAIQKSLKDPNNFRFPYKKS-----IQSEAFLPQSIPEE 67  
DB 11 YDRLVLGKLNRLCDDCVNFRFPNVAETCRSNCFYNLAFCLELYLPPSLHEE 64

Search completed: September 9, 2004, 06:17:25  
Job time : 11 secs



Db 6 SEQDILKMELEQKQKXNXT--REPVSXTAKEICWEVAKSADPLVKGVPEDKNPFKEK 63

RESULT 2

Q83173 ID 083173 PRELIMINARY; PRT; 58 AA.

AC 083173

DT 01-JUN-2003 (TReMBLrel. 24, Created)

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Ribosomal protein S21.

GN RPSU OR EF2416.

OS Enterococcus faecalis (Streptococcus faecalis).

OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.

CX NCBI\_TaxId=1351.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=V583 / ATCC 700802;

RX MEDLINE=22550857; PubMed=12663927;

RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,

RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,

RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,

RA Daugherty S., Deboy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,

RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,

RA Uterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.,

RT "Role of mobile DNA in the evolution of vancomycin-resistant

RT Enterococcus faecalis."

RL Science 299:2071-2074(2003).

DR EMBL: AE016954; AA082135.1; .

DR TIGR: EF2416; .

DR GO: GO:0005622; C:intracellular; IEA.

DR GO: GO:0005840; C:ribosome; IEA.

DR GO: GO:0003735; F:structural constituent of ribosome; IEA.

DR GO: GO:0006412; P:protein biosynthesis; IEA.

DR InterPro: IPR001911; Ribosomal S21.

DR Pfam: PF01165; Ribosomal S21; 1.

DR PRINTS: PR00976; RIBOSOMAL\_S21.

DR PRODOM: PD005521; Ribosomal\_S21; 1.

DR TIGRFAMs: TIGR00030; S21P; 1.

DR PROSITE: PS01181; RIBOSOMAL\_S21; 1.

KW Ribosomal protein; Complete proteome.

SQ SEQUENCE -58 AA; 6930 MW; ECC36D8FE7499B CRC64;

QY 7 LKTRIKQYKQSVQAGTSYRNVLQAIQKSLKDPNNFRPEPVKKSIGQSSA 58

Db 12 LDALRRFRKSVSKAGT-----LQSRK--REYVEKPSVKKKSSA 51

Query Match 14.2%; Score 54; DB 16; Length 58;

Best Local Similarity 32.7%; Pred. No. 1.9e+02;

Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

RESULT 3

Q8YUT5 ID 08YUT5 PRELIMINARY; PRT; 65 AA.

AC 08YUT5

DT 01-MAR-2002 (TReMBLrel. 20, Created)

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)

DE Hypothetical protein As9028.

GN ASR9028.

OS Anabaena sp. (strain PCC 7120).

OC Plasmid PCC7120psil10n.

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

CX NCBI\_TaxId=103690;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21595285; PubMed=11759840;

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,

RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kohara M., Matsumoto M., Matsumo A., Muraki A.,

RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing

RT cyanobacterium Anabaena sp. strain PCC 7120.";

RL DNA Res. 8:205-213(2001).

DR EMBL: AP003605; BAB77514.1; .

DR PIR: AE2575; AE2575.

DR GO: GO:0046821; C:extrachromosomal DNA; IEA.

DR GO: GO:0003677; F:DNA binding; IEA.

DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro: IPR002145; HTH\_CoqS.

DR Pfam: PF01402; HTH\_4; 1.

KW Plasmid; Hypothetical protein; Complete proteome.

SQ SEQUENCE 65 AA; 7402 MW; 5A0F7DD4ECF9127A CRC64;

QY 10 KIKYKQSVQAGTSYRNVLQAIQKSLKDPNNFRPE 46

Db 12 QIQKQELAEKLCMSTBELLSASVEDLLNYPQNEFNQ 48

Query Match 14.0%; Score 53; DB 16; Length 65;

Best Local Similarity 27.0%; Pred. No. 2.8e+02;

Matches 10; Conservative 10; Mismatches 17; Indels 0; Gaps 0;

RESULT 4

Q8EDU0 ID 08EDU0 PRELIMINARY; PRT; 74 AA.

AC 08EDU0

DT 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Transcriptional regulator, Ner family.

GN SO2693.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

CX NCBI\_TaxId=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NR-1;

RX MEDLINE=22297866; PubMed=12366813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,

RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,

RA Meyer T., Tsapin J., Scott J., Beanan M., Brinkac L., Daugherty S.,

RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,

RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,

RA Vamathevan J., Waidman J., Impraim M., Lee K., Berry K., Lee C.,

RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,

RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,

RT "Genome sequence of the dissimilatory metal ion-reducing bacterium

RT Shewanella oneidensis.";

RL Nat. Biotechnol. 20:1118-1123(2002).

DR EMBL: AE015706; AA055661.1; .

DR TIGR: SO2693; .

KW Complete proteome.

SQ SEQUENCE 74 AA; 8564 MW; 9B97312E3B71B3ED CRC64;

QY 3 HRIDIKYKQSVQAGTSYRNVLQAIQKSLKDPNNFRPEPVKKSIGQSS 56

Db 9 HKADIKALER-----AGTNYEKLAEHGIGAGSTLRNLR-----FKPKCERIIAOK 56

Query Match 13.7%; Score 52; DB 16; Length 74;

Best Local Similarity 21.1%; Pred. No. 4.1e+02;

Matches 16; Conservative 17; Mismatches 25; Indels 16; Gaps 3;

RESULT 5

Q86K6 ID 086K6 PRELIMINARY; PRT; 60 AA.

AC 086K6

DT 01-JUN-2003 (TReMBLrel. 24, Created)



```
DT 01-JUN-2003 (Tremblrel. 24, last sequence update)
DE Conserved hypothetical protein.
GN PSP01573.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000.
RA Buel R., Ucardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utecherback T., Van Aken S., Feldblyum T., Gwin M.,
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Baanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae."
RT Submitted (Mar-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF016861; AF05093.1; -.
DR TIGR; PSP01573; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 60 AA; 6685 MW; 50373D26840CD7BE CRC64;

Query Match
Best Local Similarity 24.4%; Score 50.5; DB 16; Length 60;
Matches 11; Conservative 12; Mismatches 17; Indels 5; Gaps 1;

QY 11 IQKXQVQAGTSYRNVLQAAIQKSLDPSNNRREPVKYSIQE 55
: : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : :
9 VSEFEQA-----ASVDKMFRAQVQASIDPAPNPHQVAAEMRE 48

RESULT 6
ID 006495 PRELIMINARY; PRT; 45 AA.
AC 006495;
RT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, last annotation update)
DE D-alanine-D-alanine ligase (fragment).
GN DDLG.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=840;
RC MEDLINE=97294456; PubMed=9150212;
RA Chen Y., Miyata S., Makino S., Moriyama R.;
RT "Molecular characterization of a germination-specific muramidase from
RT Clostridium perfringens S40 spores and nucleotide sequence of the
RT corresponding gene."
RT U. Bacteriol. 179:3181-3187(1997).
RL EMBL; D88151; BAA20124.1; -.
DR HSSP; F07862; 110M.
DR GO; GO:0008716; F:D-alanine-D-alanine ligase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR000291; Data_119_Van.
DR Pfam; PF01820; Data_1ligas; 1.
KW Ligase.
FT NON TER.
SQ SEQUENCE 45 AA; 5042 MW; 8A5249BDC73E835 CRC64;

Query Match
Best Local Similarity 13.2%; Score 50; DB 2; Length 45;
Matches 11; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 9 FKIQKQSVQAGTSYRNVLQAAIQKSLK 38
: : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : :
12 TKNSLFPKSAGINMSPFELLDTIEKSLK 41

DT 01-JUN-2003 (Tremblrel. 15, Created)
DE 01-OCT-2000 (Tremblrel. 15, last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE P2E2.9.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Alatafi H., Bei Q., Chin C., Chiu J., Choi E., Conn L.,
RA Connay A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharisky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC P2E2 from chromosome
RT 1."
RT Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RT Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RT Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Alatafi H., Bei B., Chin C., Chiu J., Choi E.,
RA Conn L., Connay A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharisky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RT Submitted (Jul-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC069252; AAF6561.1; -.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; Zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SMO0343; Znf_C2HC; 1.
SQ SEQUENCE 67 AA; 7713 MW; 25E733E778ABAC73 CRC64;

Query Match
Best Local Similarity 13.1%; Score 49.5; DB 10; Length 67;
Matches 16; Conservative 10; Mismatches 29; Indels 9; Gaps 2;

QY 2 IHRIDLTKIKQYK-QGVQAGTSYRNVLQAAIQKSLKDPSSNNFR-----EEPVKKS 52
: : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : :
1 WHKPSKSNFVSITCSNCGQGVHKKSCRKRAVPKPKKQGRPKKQKTMDEENFTIP 60

Db 53 IQES 56
: : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : :
61 IQSS 64

RESULT 8
ID Q8CUG4 PRELIMINARY; PRT; 69 AA.
AC Q8CUG4;
RT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Mercuric ion-binding protein.
```

```
GN OB1143.
OS Oceanobacillus theyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RA MEDLINE=22220767; PubMed=12235376;
RX Takami H., Takaki Y., Uchiyama T.;
RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL: AP004586; BAC13099.1; -.
DR GO: GO:0005507; F:copper ion binding; IEA.
DR GO: GO:0005575; F:copper ion transporter activity; IEA.
DR GO: GO:0015097; F:mercury ion transporter activity; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0005625; P:copper ion transport; IEA.
DR GO: GO:0015694; P:mercury ion transport; IEA.
DR GO: GO:0030001; P:metal ion transport; IEA.
DR InterPro: IPR000428; Cu_bind.
DR InterPro: IPR006122; Cu_bind_dom.
DR InterPro: IPR006121; Heavy_metal_transp.
DR InterPro: IPR001802; HG_scavenger.
DR Pfam: PF00403; HMA_1; Metal_bind.
DR PRINTS: PRO0944; CUENPORT.
DR PRINTS: PRO0946; HQSCAVENGER.
DR TIGRFAMs: TIGR00003; TIGR00003; 1.
DR PROSITE: PS01047; HMA_1; 1.
DR PROSITE: PS0846; HMA_2; 1.
KM Complete proteome.
SQ SEQUENCE 69 AA; 7336 MW; 834DC48A76A50E92 CRC64;

Query Match 12.9%; Score 49; DB 16; Length 69;
Best Local Similarity 25.5%; Pred. No. 7.8e+02;
Matches 14; Conservative 12; Mismatches 23; Indels 6; Gaps 1;

QY 15 KQSVQAGTSYRNTLQAAIQKSLKDPNNFREPYKKSIOSEAFIPQSIPEERY 69
DB 16 KQSVQAGTSYRNTLQAAIQKSLKDPNNFREPYKKSIOSEAFIPQSIPEERY 64

RESULT 9
ID 08ZU2 PRELIMINARY; PRT; 74 AA.
AC 08ZU2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PAE2755.
GN PAE2755.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
NCBI_TaxID=13773;
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RA MEDLINE=21664397; PubMed=11792869;
RX Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller U.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum.";
RU Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
KM EMBL: AE009892; AA64415.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 74 AA; 8826 MW; AA1AFD604B5D3646 CRC64;

Query Match 12.9%; Score 49; DB 17; Length 74;
Best Local Similarity 25.0%; Pred. No. 8.5e+02;
Matches 11; Conservative 13; Mismatches 16; Indels 4; Gaps 1;
```

```
QY 12 QKYKQSVQAGTSYRNTLQAAIQKSLKDPNNFREPYKKSIOE 55
DB 11 RELKEKARIGIDIRVERALERBEIK---RREERLAKLEE 50

RESULT 10
ID 09SUH4 PRELIMINARY; PRT; 64 AA.
AC 09SUH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE At2g42820 protein.
GN AT2G42820
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Rensing C.M., Xoo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Coppenhafer G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006931; AAD21726.1; -.
PR: F84858; F84858.
SQ SEQUENCE 64 AA; 7401 MW; A8BD423FD5B099C0 CRC64;

Query Match 12.8%; Score 48.5; DB 10; Length 64;
Best Local Similarity 35.7%; Pred. No. 8.2e+02;
Matches 15; Conservative 9; Mismatches 11; Indels 7; Gaps 2;

QY 20 GAGTSY----RNVLQAA---IQKSLKDPNNFREPYKKSIO 54
DB 16 GGGTNYTDEQRVRVQWMSLDARKSVQDYVRFQWDSVEKAIK 57

RESULT 11
ID 08LKA9 PRELIMINARY; PRT; 62 AA.
AC 08LKA9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Pathogen-related protein (Fragment).
OS Musa acuminata (Banana).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;
OC Musa.
NCBI_TaxID=4641;
OX NCBI_TaxID=4641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Cavendish;
RA Regev I., Gepstein S., Khayat E.;
RT "Identification of ripening related genes by screening banana fruit
RT SSH library.";
RU Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AF516690; AA669296.1; -.
```

```

DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
FT NON-TER 1 62
FT NON-TER 62 62
SQ SEQUENCE 62 AA; 6670 MW; 1BEE4208F8A56196 CRC64;
Query Match 12.7%; Score 48; DB 10; Length 62;
Best Local Similarity 27.7%; Pred. No. 8.9e+02;
Matches 18; Conservative 10; Mismatches 25; Indels 12; Gaps 3;

QY 16 QGVQAGNSTRNVQAAIQKSLK--DPSNNFREPVKKSIOESAF--LPQSIPEERYK 70
DB 1 QQIQMGSGSYNAFLQTSLPPELLRYDPS-----VETASSGKAFATTPRGFVNEILQ 53
QY 71 MKSKP 75
DB 54 VLSGP 58

RESULT 12
Q8IBU5 PRELIMINARY; PRT; 71 AA.
AC Q8IBU5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mitochondrial ATP synthase F1, epsilon subunit, putative (BC 3.6.3.14).
GN MAL7P1.75.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N., Quail M., Barrett B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; C:mitochondrion; IEA.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR InterPro; IPR006754; P:ATP biosynthes; IEA.
DR Pfam; PF04627; ATP-synt_Eps; 1.
KM Hydrolyase. 71 AA; 8496 MW; 213B404118FF0CCA CRC64;
SQ SEQUENCE 71 AA; 8496 MW; 213B404118FF0CCA CRC64;
Query Match 12.7%; Score 48; DB 5; Length 71;
Best Local Similarity 27.1%; Pred. No. 1e+03;
Matches 13; Conservative 12; Mismatches 23; Indels 0; Gaps 0;

QY 23 TSYRVVQAAIQKSLKDPSSNFRPEPVKKSIOESAFLPQSIPEERYK 70
DB 10 TTYASEMADILKCLDYPYSDIALERSKMTIRETYIKDGKFSQELYE 57

RESULT 13
Q49197 PRELIMINARY; PRT; 72 AA.
AC Q49197;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to Swiss-Prot Accession number P32333 (Fragment).
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G-37;
RC MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.C., Bott K.F., Hutchison C.A. III;

```

```

RT "A survey of the Mycoplasma genitalium genome by using random RT sequencing.";
RL J. Bacteriol. 175:7918-7930 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=G-37;
RA Peterson S.N.;
RL "Characterization and analysis of the Mycoplasma genitalium genome.";
RL Theiss (1992), Microbiology and Immunology, University of North Carolina Medical School.
DR EMBL; U01723; AAC4319.2; -.
FT NON-TER 1 72
FT NON-TER 72 72
SQ SEQUENCE 72 AA; 8400 MW; BAA632DF7E6D290A CRC64;
Query Match 12.7%; Score 48; DB 2; Length 72;
Best Local Similarity 32.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 9; Mismatches 24; Indels 1; Gaps 1;

QY 3 HRIDKTKIQKYSYVQAGSTRNVQAAIQKSLKDPSSNFRPEPVKKSIOESAFLPQ 62
DB 14 HRIGSKTVQVYR--IAKRTIERVCQVONK-----QELVKTTLVEDVNFYS 61
QY 63 SIPERYKM 71
DB 62 LSHELLKL 70

RESULT 14
P72995 PRELIMINARY; PRT; 72 AA.
AC P72995;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein ssi2920.
GN SSI2920.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Karako T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S., Ra Shimpu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium RT Synchocystis sp. strain PCC6803. II. Sequence determination of the RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136 (1996).
DR EMBL; D90902; BAA17014.1; -.
DR PIR; S74974; S74974.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 72 AA; 8476 MW; DD429B8C081C7F4D CRC64;
Query Match 12.5%; Score 47.5; DB 16; Length 72;
Best Local Similarity 32.0%; Pred. No. 1.2e+03;
Matches 16; Conservative 9; Mismatches 24; Indels 1; Gaps 1;

QY 7 LKT-KIQKYSYVQAGSTRNVQAAIQKSLKDPSSNFRPEPVKKSIOE 55
DB 2 LKTSFQKAIRESVNLPLDDQELIDILIQRLQKRRKKLAIEIKELRQ 51

RESULT 15
Q7WY58 PRELIMINARY; PRT; 46 AA.
AC Q7WY58;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Small acid-soluble spore protein.

```

```

GN  SSPT OR BSU33340.
OS  Bacillus subtilis subsp. subtilis str. 168.
OC  Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX  NCBI_TaxID=224308;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  MEDLINE=98044033; PubMed=9384377;
RA  Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA  Azevedo V., Bertero M.G., Bessières P., Bolotin A., Borchert S.,
RA  Borriess R., Boursier L., Brans A., Braun M., Briganti S.C., Bron S.,
RA  Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA  Choi S.K., Codani U.J., Comercon I.F., Cummings N.J., Daniel R.A.,
RA  Denizot F., Devine K.M., Dusterhot A., Ehrlich S.D., Emerson P.T.,
RA  Ertlan K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA  Fritz C., Fujita Y., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA  Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA  Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA  Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA  Kuri S.B., Karamata D., Kasanara Y., Klaerr-Blanchard M., Klein C.,
RA  Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA  Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA  Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA  Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA  Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA  Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA  Prescan E., Fujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA  Rieger M., Rivolta C., Rocha E., Roche B., Rose W., Sadate Y.,
RA  Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,
RA  Setiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Solido B.,
RA  Sorokin A., Taccout E., Takagi T., Takahashi H., Takemaru K.,
RA  Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Toomoni A.,
RA  Tosato V., Uchiyama S., Vanderbol M., Vannier F., Vassartoli A.,
RA  Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA  Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA  Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT  "The complete genome sequence of the Gram-positive bacterium Bacillus
RT  subtilis."
RL  Nature 390:249-256(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=168;
RX  Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA  Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL: Z99121; CAB01467.1; -.
SQ  SEQUENCE 46 AA; 5162 MW; B8C702960210A1E1 CRC64;

```

```

Query Match      12.4%; Score 47; DB 2; Length 46;
Best Local Similarity 25.0%; Pred. No. 8.2e+02;
Matches 8; Conservative 14; Mismatches 6; Indels 4; Gaps 1;

```

```

Oy 26 RNVIAIQKSLKDPNNFREPVKKSIOESE 57
Db 16 KAVIQGA----LEDAQSALKDDPDLQEAIVQKK 43

```

```

Search completed: September 9, 2004, 06:18:11
Job time : 42 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:13:41 ; Search time 54 Seconds  
(without alignments)  
392.427 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_62\_136  
Perfect score: 379  
Sequence: 1 NIHHDXTKTKIKYKQSVQ.....SEAFIPQSIPEERYKMKSKP 75

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 806123

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp.1980s:\*
- 2: geneseqp.1990s:\*
- 3: geneseqp.2000s:\*
- 4: geneseqp.2001s:\*
- 5: geneseqp.2002s:\*
- 6: geneseqp.2003as:\*
- 7: geneseqp.2003bs:\*
- 8: geneseqp.2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.5	14.6	75	4	AAM90241 Human imm
2	54	14.2	56	4	AU035471 Enterococ
3	54	14.2	58	4	AU035472 Enterococ
4	54	14.2	58	6	ABU14651 Protein e
5	53	14.0	58	6	ABU46553 Protein e
6	53	14.0	66	6	ABU4678 Protein e
7	53	14.0	70	5	ABP27318 Streptoco
8	53	14.0	70	5	ABP30108 Streptoco
9	53	14.0	72	5	ABP30812 Streptoco
10	52.5	13.9	59	4	ABP10556 Human pan
11	52.5	13.9	59	4	AAM92092 Human dig
12	52	13.7	71	5	ABP3328 Human CRI
13	51.5	13.6	66	3	AAE14202 Human Lyr
14	51.5	13.6	66	5	AAE14628 Human Lyr
15	51	13.5	53	4	AAM20732 Peptide #
16	51	13.5	53	4	ABP42197 Peptide #
17	51	13.5	53	4	AAM36004 Peptide #
18	51	13.5	53	4	ABP25745 Peptide #
19	51	13.5	53	4	AAM75893 Human bon
20	51	13.5	53	4	AAM63082 Human bra
21	51	13.5	53	4	ABG57629 Human liv
22	51	13.5	53	5	ABG45336 Human pep
23	51	13.5	58	4	AAM16656 Peptide #
24	51	13.5	58	4	AAM15183 Peptide #
25	51	13.5	58	4	ABP35640 Peptide #

26	51	13.5	58	4	ABP34176	Abp34176 Peptide #
27	51	13.5	58	4	AAM27642	Aam27642 Peptide #
28	51	13.5	58	4	AAM29140	Aam29140 Peptide #
29	51	13.5	58	4	ABP30473	Abp30473 Peptide #
30	51	13.5	58	4	ABP29009	Abp29009 Peptide #
31	51	13.5	58	4	ABP19617	Abp19617 Protein #
32	51	13.5	58	4	ABP21069	Abp21069 Protein #
33	51	13.5	58	4	AAM68834	Aam68834 Human bon
34	51	13.5	58	4	AAM67350	Aam67350 Human bon
35	51	13.5	58	4	AAM54967	Aam54967 Human bra
36	51	13.5	58	4	AAM56456	Aam56456 Human bra
37	51	13.5	58	4	ABG50494	Abg50494 Human liv
38	51	13.5	58	4	ABG49013	Abg49013 Human liv
39	51	13.5	58	4	AAM02922	Aam02922 Peptide #
40	51	13.5	58	4	AAM04372	Aam04372 Peptide #
41	51	13.5	58	4	ABG38414	Abg38414 Human pep
42	51	13.5	58	5	ABG36985	Abg36985 Human pep
43	51	13.5	65	7	ADG96014	Adg96014 E. faeciu
44	50	13.2	57	5	ABP49622	Abp49622 Listeria
45	50	13.2	57	6	ABU32772	Abu32772 Protein e

## ALIGNMENTS

RESULT 1	
AAM90241	
ID	AAM90241 standard; protein; 75 AA.
XX	
AC	AAM90241;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen SEQ ID NO:17834.
XX	
XX	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM	Cytotoxic; gene therapy; vaccine; metastasis.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US001354.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180528P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-019874P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-020515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	07-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	11-JUL-2000; 2000US-0217487P.
PR	14-JUL-2000; 2000US-0218290P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225447P.

	PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
	PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
	PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
	PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
	PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
	PR	22-AUG-2000;	2000US-0226688P.	PR	17-NOV-2000;	2000US-0249208P.
	PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
	PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
	PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
	PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
	PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
	PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
	PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
	PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
	PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
	PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
	PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
	PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249246P.
	PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249245P.
	PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249297P.
	PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249299P.
	PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
	PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0250160P.
	PR	14-SEP-2000;	2000US-0232397P.	PR	01-DEC-2000;	2000US-0250391P.
	PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251030P.
	PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0251988P.
	PR	14-SEP-2000;	2000US-0232400P.	PR	05-DEC-2000;	2000US-0256719P.
	PR	14-SEP-2000;	2000US-0232401P.	PR	06-DEC-2000;	2000US-0251479P.
	PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251856P.
	PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251868P.
	PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251869P.
	PR	21-SEP-2000;	2000US-0234232P.	PR	08-DEC-2000;	2000US-0251989P.
	PR	21-SEP-2000;	2000US-0234374P.	PR	11-DEC-2000;	2000US-0251990P.
	PR	23-SEP-2000;	2000US-0234597P.	PR	05-JAN-2001;	2001US-0256768P.
	PR	25-SEP-2000;	2000US-0234988P.		(HUMA-) HUMAN GENOME SCI INC.	
	PR	26-SEP-2000;	2000US-0235484P.	PA	Rosen CA, Barash SC, Ruben SW,	
	PR	27-SEP-2000;	2000US-0235834P.	PI	WPT; 2001-463426/52.	
	PR	29-SEP-2000;	2000US-0235836P.	XX	N-PsDB; AAK63022.	
	PR	29-SEP-2000;	2000US-0236327P.	DR	WT; 2001-463426/52.	
	PR	29-SEP-2000;	2000US-0236368P.	XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
	PR	29-SEP-2000;	2000US-0236369P.	PT	useful for preventing, diagnosing and/or treating cancers and metastasis.	
	PR	02-OCT-2000;	2000US-0237037P.	XX	Claim 11; SEQ ID NO 17834; 3071pp + Sequence listing; English.	
	PR	02-OCT-2000;	2000US-0237038P.	PS	AAK5951 to AAK64702 encode the human immune/hematopoietic antigen (I)	
	PR	02-OCT-2000;	2000US-0237040P.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
	PR	13-OCT-2000;	2000US-0239935P.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
	PR	20-OCT-2000;	2000US-0240960P.	CC	treatment of diseases associated with inappropriate (I) expression. For	
	PR	20-OCT-2000;	2000US-0241221P.	CC	example, they may be used to treat disorders associated with decreased	
	PR	20-OCT-2000;	2000US-0241785P.	CC	expression by rectifying mutations or deletions in a patient's genome	
	PR	20-OCT-2000;	2000US-0241786P.	CC	that affect the activity of (I) by expressing	



Db 12 LDDALRRFKSVSKAGT-----LQESRK---REFYKPSVKKKKSEA 51

```
RESULT 4
ABU14651
ID ABU14651 standard; protein; 58 AA.
XX
XX ABU14651;
AC
XX 19-JUN-2003 (first entry)
DT
XX Protein encoded by Prokaryotic essential gene #178.
DE
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Enterococcus faecalis.
OS
XX WO200277183-A2.
PN
XX 03-OCT-2002.
PD
XX 21-MAR-2002; 2002WO-US009107.
PF
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
DR N-PSDB; ACA18521.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 42575; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation; (7) identifying a gene in an operon required for
CC proliferation; (8) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by this
CC parent did not form part of the printed specification, but was obtained
CC in electronic format directly from WMO at
```

CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 58 AA;

Query Match 14.2%; Score 54; DB 6; Length 58;  
Best Local Similarity 32.7%; Pred. No. 1.1e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

```
QY 7 LKTKIQKYGQVQAGTGYRVLQALQKSLKDPSSNFFREBPVKQIQSEEA 58
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 LDDALRRFKSVSKAGT-----LQESRK---REFYKPSVKKKKSEA 51
```

```
RESULT 5
ABU46553
ID ABU46553 standard; protein; 58 AA.
XX
XX ABU46553;
AC
XX 19-JUN-2003 (first entry)
DT
XX Protein encoded by Prokaryotic essential gene #32080.
DE
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Streptococcus pyogenes.
OS
XX WO200277183-A2.
PN
XX 03-OCT-2002.
PD
XX 21-MAR-2002; 2002WO-US009107.
PF
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
DR N-PSDB; ACA50423.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 74477; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation; (7) identifying a gene in an operon required for
CC proliferation; (8) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
```



CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 58 AA;  
Query Match 14.0%; Score 53; DB 6; Length 58;  
Best Local Similarity 32.7%; Pred. No. 1.4e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;  
QY 7 LKTKIQKXKQSGAGTSYRNVLQALQKSLKDPSSNNFREBPVKSIQSEEA 58  
DB 12 LDDALRRFKRSVTKAGT-----LQESRK---REFYKPSVKRKKESEA 51  
RESULT 6  
ABU44678  
ID ABU44678 standard; protein; 66 AA.  
XX  
AC ABU44678;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #30205.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Streptococcus mutans.  
XX  
WO20027183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0352699P.  
XX  
PA (ELITR-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
XX  
N-PSDB; ACA48548.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 72602; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound acts; (10) profiling a  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 66 AA;  
Query Match 14.0%; Score 53; DB 6; Length 66;  
Best Local Similarity 32.7%; Pred. No. 1.6e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;  
QY 7 LKTKIQKXKQSGAGTSYRNVLQALQKSLKDPSSNNFREBPVKSIQSEEA 58  
DB 20 LDDALRRFKRSVTKAGT-----LQESRK---REFYKPSVKRKKESEA 59  
RESULT 7  
ABP27318  
ID ABP27318 standard; protein; 70 AA.  
XX  
AC ABP27318;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Streptococcus polypeptide SEQ ID NO 3812.  
XX  
KM Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX antinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus pyogenes.  
XX  
WO200234771-A2.  
XX  
PD 02-MAY-2002.  
XX  
PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Frazer C;  
PI Tettelein H;  
XX  
DR WPI; 2002-352536/38.  
XX  
N-PSDB; ABN67949.  
XX  
PT New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
PS Claim 1; Page 3536; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 70 AA;

Query Match 14.0%; Score 53; DB 5; Length 70;  
Best Local Similarity 32.7%; Pred. No. 1.7e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

Db 7 LKTKIKYKQSVGAGTSYRNVLQAIQKSLKDPNNFRPEPVKKSIOESSEA 58  
24 LDDALRRKRSYVTKAGT-----LQESRK--REFYKPSVKRKKESEA 63

## RESULT 8

ABP30108  
ID ABP30108 standard; protein; 70 AA.

XX  
AC ABP30108;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 9392.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KM group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

XX  
FN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;

XX Tettelein H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN70739.

XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX Claim 1; Page 4064; 4525BP; English.  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 70 AA;

Query Match 14.0%; Score 53; DB 5; Length 70;  
Best Local Similarity 32.7%; Pred. No. 1.7e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

Db 7 LKTKIKYKQSVGAGTSYRNVLQAIQKSLKDPNNFRPEPVKKSIOESSEA 58  
24 LDDALRRKRSYVTKAGT-----LQESRK--REFYKPSVKRKKESEA 63

## RESULT 9

ABP30812  
ID ABP30812 standard; protein; 72 AA.

XX  
AC ABP30812;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 10800.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KM group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

XX  
FN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

XX 07-MAR-2001; 2001GB-00005640.

PA (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;

XX Tettelein H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN71443.

XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX Claim 1; Page 4185; 4525BP; English.  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

Query Match	14.0%;	Score 53;	DB 5;	Length 72;
Best Local Similarity	32.7%;	Pred. No. 1.8e+02;		
Matches 17;	Conservative 10;	Mismatches 13;	Indels 12;	Gaps 2
Qy	7	LKTKIOKYSVOGAGTSYRVNVLQALIKSLKDPSSNFRFEPVKKSIOSEA	58	
Db	26	LDALRRFRGSRVTKAGT-----LQSKR--REFEYKPSVKRRKSEA	65	
RESULT 10				
ABBI0556				
ID	ABBI0556	standard; protein; 59 AA.		
XX	AC	ABBI0556;		
XX	DT	14-JAN-2002 (first entry)		
XX	DE	Human pancreatic cancer related polypeptide, SEQ ID NO: 205.		
XX	KW	Human; cytotstatic; antidiabetic; antiinflammatory; gastric; osteopathic; antihormone; antilacer; thyroid-active; gene therapy; antisense therapy; pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis; diabetes; endocrine disorder; acromegaly; hyperthyroidism; gastrointestinal disorder; Crohn's disease; duodenal ulcer.		
XX	OS	Homo sapiens.		
XX	PN	WO200155206-A1.		
XX	PD	02-AUG-2001.		
XX	PF	17-JAN-2001; 2001WO-US001353.		
XX	PR	31-JAN-2000; 2000US-0179065P.		
XX	PR	04-FEB-2000; 2000US-0180628P.		
XX	PR	24-FEB-2000; 2000US-0184664P.		
XX	PR	02-MAR-2000; 2000US-0186350P.		
XX	PR	16-MAR-2000; 2000US-0189874P.		
XX	PR	17-MAR-2000; 2000US-0190076P.		
XX	PR	18-APR-2000; 2000US-0198123P.		
XX	PR	19-MAY-2000; 2000US-0205515P.		
XX	PR	07-JUN-2000; 2000US-0209467P.		
XX	PR	30-JUN-2000; 2000US-0214886P.		
XX	PR	28-JUN-2000; 2000US-0215135P.		
XX	PR	07-JUL-2000; 2000US-0216647P.		
XX	PR	07-JUL-2000; 2000US-0216880P.		
XX	PR	11-JUL-2000; 2000US-0217487P.		
XX	PR	11-JUL-2000; 2000US-0217496P.		
XX	PR	14-JUL-2000; 2000US-0218290P.		
XX	PR	26-JUL-2000; 2000US-0220663P.		
XX	PR	26-JUL-2000; 2000US-0220964P.		
XX	PR	14-AUG-2000; 2000US-0224518P.		
XX	PR	14-AUG-2000; 2000US-0224519P.		
XX	PR	14-AUG-2000; 2000US-0225213P.		
XX	PR	14-AUG-2000; 2000US-0225214P.		
XX	PR	14-AUG-2000; 2000US-0225266P.		
XX	PR	14-AUG-2000; 2000US-0225267P.		
XX	PR	14-AUG-2000; 2000US-0225268P.		

PR	14-AUG-2000	2000US-02254770
PR	14-AUG-2000	2000US-02254476
PR	14-AUG-2000	2000US-02267577
PR	14-AUG-2000	2000US-02267588
PR	14-AUG-2000	2000US-02267599
PR	18-AUG-2000	2000US-02262671
PR	18-AUG-2000	2000US-02262679
PR	22-AUG-2000	2000US-02266689
PR	22-AUG-2000	2000US-02271822
PR	23-AUG-2000	2000US-02271009
PR	30-SEP-2000	2000US-02282944
PR	01-SEP-2000	2000US-02282878
PR	01-SEP-2000	2000US-02293434
PR	01-SEP-2000	2000US-02293452
PR	05-SEP-2000	2000US-02292539
PR	05-SEP-2000	2000US-02285159
PR	06-SEP-2000	2000US-02304379
PR	06-SEP-2000	2000US-02304382
PR	08-SEP-2000	2000US-02312422
PR	08-SEP-2000	2000US-02312443
PR	08-SEP-2000	2000US-02314132
PR	08-SEP-2000	2000US-02314149
PR	08-SEP-2000	2000US-02307060
PR	08-SEP-2000	2000US-02330814
PR	12-SEP-2000	2000US-02331968
PR	14-SEP-2000	2000US-02333397
PR	14-SEP-2000	2000US-02333398
PR	14-SEP-2000	2000US-02323399
PR	14-SEP-2000	2000US-02334000
PR	14-SEP-2000	2000US-02334010
PR	14-SEP-2000	2000US-02330633
PR	14-SEP-2000	2000US-02330644
PR	14-SEP-2000	2000US-02330656
PR	21-SEP-2000	2000US-02342233
PR	21-SEP-2000	2000US-02342274
PR	25-SEP-2000	2000US-02349376
PR	25-SEP-2000	2000US-02349398
PR	26-SEP-2000	2000US-02354844
PR	27-SEP-2000	2000US-02358344
PR	27-SEP-2000	2000US-02358356
PR	28-SEP-2000	2000US-02363327
PR	29-SEP-2000	2000US-02363676
PR	29-SEP-2000	2000US-02363689
PR	29-SEP-2000	2000US-02363699
PR	02-OCT-2000	2000US-02363702
PR	02-OCT-2000	2000US-02368029
PR	02-OCT-2000	2000US-02370137
PR	02-OCT-2000	2000US-02370389
PR	02-OCT-2000	2000US-02370398
PR	02-OCT-2000	2000US-02370400
PR	13-OCT-2000	2000US-02393935
PR	13-OCT-2000	2000US-02393937
PR	20-OCT-2000	2000US-02409660
PR	20-OCT-2000	2000US-02412122
PR	20-OCT-2000	2000US-02417858
PR	20-OCT-2000	2000US-02417866
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02418088
PR	20-OCT-2000	2000US-02418093
PR	20-OCT-2000	2000US-02418262
PR	01-NOV-2000	2000US-02446187
PR	08-NOV-2000	2000US-02464747
PR	08-NOV-2000	2000US-02464752
PR	08-NOV-2000	2000US-02464765
PR	08-NOV-2000	2000US-02464777
PR	08-NOV-2000	2000US-02464789
PR	08-NOV-2000	2000US-02465233
PR	08-NOV-2000	2000US-02465244
PR	08-NOV-2000	2000US-02465252
PR	08-NOV-2000	2000US-02465265
PR	08-NOV-2000	2000US-02465274









*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:17:18 ; Search time 48 Seconds

(without alignments)  
501.078 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_62\_136

Perfect score: 379  
Sequence: 1 NIHRLDKTKIKYKQGVQ.....SEAFPOSIREPKYKSKRP 75

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 133516 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 467853

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	15.6	73	16 US-10-767-701-55558	Sequence 55558, A
2	54	14.2	56	9 US-09-815-242-4967	Sequence 4967, Ap
3	54	14.2	56	9 US-09-815-242-18815	Sequence 18815, A
4	54	14.2	58	12 US-10-282-122A-42575	Sequence 42575, A
5	53.5	14.1	58	16 US-10-767-701-41772	Sequence 41772, A
6	53	14.0	58	12 US-10-282-122A-74477	Sequence 74477, A
7	53	14.0	58	12 US-10-282-122A-72602	Sequence 72602, A
8	52.5	13.9	55	12 US-10-424-599-188405	Sequence 188405, A
9	52	13.7	71	11 US-09-864-408A-2602	Sequence 2602, Ap
10	51	13.5	53	9 US-09-864-761-41043	Sequence 41043, A
11	51	13.5	58	9 US-09-864-761-34915	Sequence 34915, A
12	51	13.5	58	9 US-09-864-761-36367	Sequence 36367, A
13	50.5	13.3	46	12 US-10-424-599-157322	Sequence 157322, A
14	50	13.2	57	12 US-10-282-122A-60696	Sequence 60696, A
15	50	13.2	60	12 US-10-424-599-216377	Sequence 216377, A

16	50	13.2	64	12 US-10-424-599-189867	Sequence 189867, A
17	49.5	13.1	63	12 US-10-424-599-166391	Sequence 166391, A
18	49.5	13.1	67	9 US-09-864-761-36261	Sequence 36261, A
19	49	12.9	44	10 US-09-845-917-39	Sequence 39, Appl
20	49	12.9	44	10 US-09-845-917A-39	Sequence 39, Appl
21	49	12.9	51	9 US-09-864-761-47571	Sequence 47571, A
22	49	12.9	58	9 US-08-815-242-12568	Sequence 12568, A
23	49	12.9	58	9 US-09-815-242-12969	Sequence 12969, A
24	49	12.9	58	9 US-09-908-931B-52	Sequence 52, Appl
25	49	12.9	58	12 US-10-282-122A-70357	Sequence 70357, A
26	49	12.9	58	12 US-10-282-122A-71237	Sequence 71237, A
27	49	12.9	58	12 US-10-282-122A-71760	Sequence 71760, A
28	49	12.9	58	12 US-10-332-964-52	Sequence 52, Appl
29	49	12.9	64	12 US-10-424-599-26008	Sequence 26008, A
30	49	12.9	68	16 US-10-437-963-204780	Sequence 204780, A
31	49	12.9	70	12 US-10-424-599-228746	Sequence 228746, A
32	49	12.9	74	12 US-10-425-114-49589	Sequence 49589, A
33	48.5	12.8	73	12 US-10-424-599-193168	Sequence 193168, A
34	48	12.7	44	9 US-09-864-761-34364	Sequence 34364, A
35	48	12.7	45	9 US-09-864-761-37578	Sequence 37578, A
36	48	12.7	63	9 US-08-864-761-44034	Sequence 44034, A
37	48	12.7	74	9 US-09-864-761-47086	Sequence 47086, A
38	47.5	12.5	54	12 US-10-424-599-218803	Sequence 218803, A
39	47.5	12.5	61	14 US-10-083-357-933	Sequence 933, App
40	47.5	12.5	74	12 US-10-424-599-258622	Sequence 258622, A
41	47	12.4	55	12 US-10-424-599-271375	Sequence 271375, A
42	47	12.4	58	16 US-10-437-963-117939	Sequence 117939, A
43	47	12.4	60	9 US-09-864-761-43072	Sequence 43072, A
44	47	12.4	60	14 US-10-029-386-30969	Sequence 30969, A
45	47	12.4	68	10 US-09-866-066-36	Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-10-767-701-55558  
; Sequence 55558, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 55558  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 30164061.pep  
US-10-767-701-55558

Query Match 15.6%, Score 59; DB 16; Length 73;  
Best Local Similarity 30.2%; Pred. No. 34;  
Matches 16; Conservative 12; Mismatches 15; Indels 10; Gaps 2;  
QY 2 NIHRLDKTKIKYKQGVQAGTSYRNLQAAIQSLDPSNNFREPKYSIQ 54  
DB 29 INDTKSKGQNYK-----YSDLFRK--QKEMSEINNFKEIKKIQ 71  
RESULT 2  
US-09-815-242-4967  
; Sequence 4967, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.



Db 12 LDDALRRFKRSVTKAGT-----LQESRK---REFYKPSVKRKKESEA 51

RESULT 5

US-10-767-701-41772

Sequence 41772, Application US/10767701  
Publication No. US20040172684A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongmei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(5353)B  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 41772

LENGTH: 63

TYPE: PRT

ORGANISM: Sorghum bicolor

FEATURE:

NAME/KEY: unsure

LOCATION: (1)-(63)

OTHER INFORMATION: unsure at all Xaa locations

FEATURE: OTHER INFORMATION: Clone ID: LIB3476-021-P1-K1-B8-pep

US-10-767-701-41772

Query Match 14.1% Score 53.5; DB 16; Length 63;

Best Local Similarity 40.0%; Pred. No. 1.2e+02;

Matches 16; Conservative 4; Mismatches 19; Indels 1; Gaps 1;

QY 35 KSLKDPNNFREPEPVKKSIOESSEALPDSIPEERYKXKSK 74

Db 25 KDLKKENAEFRERAKSLISINE-FLKPAEGERYKXSSR 63

RESULT 6

US-10-282-122A-74477

Sequence 74477, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 74477  
LENGTH: 58  
TYPE: PRT  
ORGANISM: Streptococcus pyogenes

US-10-282-122A-74477

Query Match 14.0% Score 53; DB 12; Length 58;

Best Local Similarity 32.7%; Pred. No. 1.2e+02;

Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

QY 7 LTKIKQYKQSVGAGTSYRNVLQAIQKSLKDPNNFREPEPVKKSIOESSE 58

Db 12 LDDALRRFKRSVTKAGT-----LQESRK---REFYKPSVKRKKESEA 51

RESULT 7

US-10-282-122A-72602

Sequence 72602, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 72602

LENGTH: 66

TYPE: PRT

ORGANISM: Streptococcus mutans

US-10-282-122A-72602

Query Match 14.0%; Score 53; DB 12; Length 66;  
Best Local Similarity 32.7%; Pred. No. 1.5e+02;  
Matches 17; Conservative 10; Mismatches 13; Indels 12; Gaps 2;

Qy 7 LKTIQKXKQSVQAGTSYRNVLQAAIQKSLDPSNNFPREPVKKSIOGSEA 58  
Db 20 LDDALRRFRKRSYTKAGT-----LQESRK---REFYKRPVKRKRKSEA 59

RESULT 8  
US-10-424-599-188405  
Sequence 188405, Application US/10424599  
Publication No. US20040031072A1

GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ. ID NOS: 285684  
SEQ. ID NO: 188405  
LENGTH: 55  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_141144C.1 pep

US-10-424-599-188405

Query Match 13.9%; Score 52.5; DB 12; Length 55;  
Best Local Similarity 28.6%; Pred. No. 1.3e+02;  
Matches 12; Conservative 7; Mismatches 22; Indels 1; Gaps 1;

Qy 3 HRIDLKTIQKXKQSVQAGTSYRNVLQAAIQKSLDPSNNF 44  
Db 13 HVSLSLTKTSKXNTTAMSSPRYSICMTTTHSL-DETNVY 53

RESULT 9

US-09-864-408A-2602  
Sequence 2602, Application US/09864408A  
Publication No. US20040009474A1

GENERAL INFORMATION:  
APPLICANT: Leach, Martin D.  
APPLICANT: Shinkets, Richard A.  
TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc  
FILE REFERENCE: 21402-012  
CURRENT APPLICATION NUMBER: US/09/864,408A  
CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: 60/206,690  
PRIOR FILING DATE: 2000-05-24  
NUMBER OF SEQ. ID NOS: 9068  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ. ID NO: 2602  
LENGTH: 71  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-408A-2602

Query Match 13.7%; Score 52; DB 11; Length 71;  
Best Local Similarity 28.1%; Pred. No. 2.1e+02;  
Matches 16; Conservative 9; Mismatches 16; Indels 16; Gaps 2;

Qy 9 TKIQKXKQSVQAGTSYRNVLQAAIQKSLDPSNNFPREPVKKSIOGSEARLPQSI 65  
Db 20 TKVEDLSFEVQAIHCHN-----SWKELKKEPRNTTPOEP-----PQSLP 60

RESULT 10

US-09-864-761-41043  
Sequence 41043, Application US/09864761  
Patent No. US20020048763A1

GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ. ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ. ID NO: 41043  
LENGTH: 53  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL121897.15  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2  
OTHER INFORMATION: EST HUMAN HIT: BE866833.1, EVALUATE 4.00e-27  
OTHER INFORMATION: SWISSPROT HIT: P32802, EVALUATE 5.00e-04

US-09-864-761-41043

Query Match 13.5%; Score 51; DB 9; Length 53;  
Best Local Similarity 32.1%; Pred. No. 1.5e+02;  
Matches 9; Conservative 10; Mismatches 7; Indels 2; Gaps 1;



```

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EST HUMAN HIT: AM372226.1, EVALUE 1.00e-30
; OTHER INFORMATION: SWISSPROT HIT: P32802, EVALUE 2.00e-05
; US-09-864-761-36367

Query Match      13.5%; Score 51; DB 9; Length 58;
Best Local Similarity 32.1%; Pred. No. 2.1e+02;
Matches 9; Conservative 10; Mismatches 7; Indels 2; Gaps 1;

OY      44  FREDPVKKSIOSEAFLOPSIPPEYKYM 71
DB      29  FRKOPYDNPVRTNQ--IPROIPQPRQWYM 54

RESULT 13
US-10-424-599-157322
; Sequence 157322, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 285684
; SEQ ID NO 157322
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113082C.1.pep
; US-10-424-599-157322

Query Match      13.3%; Score 50.5; DB 12; Length 46;
Best Local Similarity 34.5%; Pred. No. 1.8e+02;
Matches 10; Conservative 9; Mismatches 7; Indels 3; Gaps 1;

OY      46  EEPVKKSIQSEAFLOPSIPPEYKYSK 74
DB      21  EDPILISQLEKQEHV--SSPRKLYKIKK 46

RESULT 14
US-10-282-122A-60696
; Sequence 60696, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EMBLRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60696
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
; US-10-282-122A-60696

Query Match      13.2%; Score 50; DB 12; Length 57;
Best Local Similarity 28.8%; Pred. No. 2.7e+02;
Matches 15; Conservative 13; Mismatches 12; Indels 12; Gaps 2;

OY      7  LKTKIQKXKOSYOGAGTSYRNVLQAIQKSLKPSNFRPEPVKKSIOSEEA 58
DB      12  LEDALRRKRTYSKSGT-----LQESRK--REFYKPSVKRKKESEA 51

RESULT 15
US-10-424-599-216377
; Sequence 216377, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 216377
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_37417C.1.pep
; US-10-424-599-216377

Query Match      13.2%; Score 50; DB 12; Length 60;
Best Local Similarity 31.2%; Pred. No. 2.9e+02;
Matches 15; Conservative 10; Mismatches 9; Indels 14; Gaps 3;

OY      24  STRNVLOAI-----QKSIKPSNNFR-EEPVKKSIOSEAFLOPSIPB 66
DB      22  SYKNNIKASLPSNKQKGVKNIENFTLSNPLKR-----PFSAPK 60

Search completed: September 9, 2004, 06:19:53
```

Job time : 51 secs

---

This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:16:18 / Search time 20 Seconds  
(without alignments)  
193.597 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_62\_136

Sequence: 1 NHRIDLTKTKIQYKQSVQ.....SEAFPLPSIPERYKMKRP 75

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 252506

Minimum DB seq length: 0  
Maximum DB seq length: 75

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/2/1aa/PCTOS\_COMB.pep:\*

6: /cgn2\_6/ptodata/2/1aa/backfiletest.pep:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	51.5	13.6	66 3	US-09-020-116-1
2	51.5	13.6	66 4	US-09-608-902-1
3	51	13.5	65 4	US-09-107-532A-5641
4	50	13.2	73 1	US-08-280-443-25
5	50	13.2	73 1	US-08-457-459-25
6	50	13.2	73 1	US-08-555-678-25
7	50	13.2	73 5	PCT-US85-02275-25
8	49.5	13.1	40 3	US-09-020-116-5
9	49.5	13.1	40 4	US-09-608-902-5
10	49	12.9	44 4	US-09-845-917A-39
11	49	12.9	44 4	US-09-732-210-1542
12	47.5	12.5	67 3	US-09-045-631-12
13	47.5	12.5	67 3	US-09-158-843A-12
14	45	11.9	31 4	US-09-187-788-62
15	45	11.9	31 4	US-09-139-600-57
16	45	11.9	37 3	US-09-302-596-11
17	45	11.9	37 3	US-09-333-415-11
18	45	11.9	37 4	US-09-303-016-11
19	45	11.9	37 4	US-09-805-507-11
20	45	11.9	38 1	US-08-519-180-9
21	45	11.9	38 1	US-09-302-596-10
22	45	11.9	38 4	US-09-333-415-10
23	45	11.9	38 4	US-09-303-016-10
24	45	11.9	38 4	US-09-805-507-10
25	45	11.9	62 4	US-09-107-532A-6344
26	45	11.9	65 4	US-09-621-976-6681
27	44	11.6	48 1	US-08-546-540-100

28	44	11.6	48 5	PCT-US96-09809-100	Sequence 100, App
29	44	11.6	69 4	US-09-134-001C-3885	Sequence 3885, Ap
30	43.5	11.5	74 3	US-08-613-822-20	Sequence 20, Appl
31	43.5	11.5	74 4	US-09-479-7298-20	Sequence 20, Appl
32	43.5	11.5	74 4	US-09-366-887A-25	Sequence 25, Appl
33	43.5	11.5	74 4	US-09-717-209-20	Sequence 20, Appl
34	43	11.3	38 4	US-08-082-315-22	Sequence 22, Appl
35	43	11.3	38 4	US-09-733-524A-22	Sequence 22, Appl
36	43	11.3	70 4	US-09-620-956-33	Sequence 33, Appl
37	43	11.3	70 4	US-09-611-152-33	Sequence 33, Appl
38	43	11.3	70 4	US-09-631-531-33	Sequence 33, Appl
39	42.5	11.2	43 3	US-09-156-580-12	Sequence 12, Appl
40	42.5	11.2	43 3	US-09-156-579C-14	Sequence 14, Appl
41	42.5	11.2	43 4	US-09-015-030-7	Sequence 7, Appl
42	42.5	11.2	68 4	US-09-328-352-7346	Sequence 7346, Ap
43	42.5	11.2	68 4	US-09-328-352-7697	Sequence 7697, Ap
44	42.5	11.2	68 4	US-09-107-532A-6707	Sequence 6707, Ap
45	42	11.1	42 3	US-08-956-307B-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-020-116-1  
Sequence 1, Application US/09020116  
Patent No. 6084063  
GENERAL INFORMATION:  
APPLICANT: Vonakis, Becky  
APPLICANT: Metzger, Henry  
TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS  
TITLE OR INVENTION: OF ALLERGIC REACTIONS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESS: NEEDLE & ROSENBERG, P.C.  
STREET: SUITE 1200, 127 PEACHTREE STREET  
CITY: ATLANTA  
STATE: GA  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/020,116  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: MILLER, MARY L  
REGISTRATION NUMBER: 39,303  
REFERENCE/DOCKET NUMBER: 14014,0285  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-020-116-1  
Query Match 13.6%, Score 51.5, DB 3, Length 66,  
Best Local Similarity 23.5%, Prod. No. 39,  
Matches 16, Conservative 14, Mismatches 13, Indels 25, Gaps 3;

```
OY      5 ILLKRIQIKVQAGTSYRNVLQAALGKSLKDPSNNFREPYPKSKTGSSEAFILPSTI 64
           :|::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      17 VOLKI-----QPVNTERITLY-----VRDPTSMKKQRPFVS-----QLL 51
OY      65 PEERYMK 72
Db      52 PGARFOTK 59
```

```

RESULT 2
US-09-608-902-1
; Sequence 1, Application US/09608902
; Patent No. 6353097
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF
; TITLE OF INVENTION: ALLERGIC REACTIONS
; FILE REFERENCE: 14014.028502
; CURRENT APPLICATION NUMBER: US/09/608.902
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/020.116
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 66
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic consen
; US-09-608-902-1

```

```

Query Match 13.6% Score 51.5; DB 4; Length 66;
Best Local Similarity 23.5%; Pred. No. 39;
Matches 16; Conservative 14; Mismatches 13; Indels 25; Gaps 3

QY 5 ILKTKIKQIKKRSVCGAGTSYRVYQALIKSLDPSNNFRBRYKKSLDSEAFIPQSI 64
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 VDKIT-----QPVNTEKTI-----VADPSNKKQKRVPS-----QLL 51
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 PEERYKMK 72
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 PGQRFQTK 59
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 3  
US-09-107-532A-5641  
; Sequence 5641 Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 6C/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Atinello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5641:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 65 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEetical: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (B) LOCATION 1..65  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5641:  
 IS-09-107-532A-5641

Query Match 13.5%; Score 51; DB 4; Length 65;  
Best Local Similarity 30.8%; Pred. No. 44;  
Matches 16; Conservative 11; Mismatches 13; Gaps 2;

```

0y      7 LKTIKIKYKOSVQGAGTSYRNVTQAIIQKSLDPSNNFREEPYKKSIOESEA 58
      ::::| | | | : | | | : | | |
Db     19 LDDALRRFKRSVSKAGT-----LQESRK---REFYEKPSVRKKKSEA 58

```

RESULT 4  
 US-08-280-443-25  
 ; Sequence 25; Application US/08280443  
 ; Patent No. 5643778  
 ; GENERAL INFORMATION:  
 APPLICANT: Nishikura, Kazuko  
 TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
 TITLE OF INVENTION: Thereof  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Howson and Howson  
 STREET: Spring House Corporate Cntr, P.O. Box 457  
 City: Spring House  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19477  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 \* OPERATING SYSTEM: PC-DOS/MS-DOS  
 \* SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/280,443  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/197,794  
 FILING DATE: 17-FEB-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bak, Mary E  
 REGISTRATION NUMBER: 31,215  
 REFERENCE/DOCKET NUMBER: W5749AUSA  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-540-9206  
 TELEFAX: 215-540-5818  
 INFORMATION FOR SEQ ID NO: 25:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 73 amino acids  
 TYPE: amino acid

```

      ; TOPOLOGY: unknown
      ; MOLECULE TYPE: protein
      US-08-280-443-25

Query Match      13.2% ; Score 50 ; DB 1 ; Length 73 ;
Best Local Similarity 23.9% ; Pred. No. 66 ;
Matches 16 ; Conservative 7 ; Mismatches 10 ; Indels 34 ; Gaps 1 ;

CY      6 DLRKTRIQXKOS-----VQAGTSYRNVLCQ 31
      |||:::|
Db      2 DPKTRIQEVLQGHRLPFTYLVVQVRGAHDQETTHCQVSGLSPEVVGTSRRKAQCA 61
      |||:::|

      32 AIQKSLK 38
      |::|
Db      62 AAEQALK 68

RESULT 5
US-08-457-459-25
; Sequence 25, Application US/08457459
; Patent No. 5677428
GENERAL INFORMATION:
APPLICANT: Nishikura, Kazuko
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
TITLE OF INVENTION: Thecod
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESSES:
ADDRESSES: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,459
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/280,443
FILING DATE: 25-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WSI49CUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 amino acids
TYPE: amino acid
TOPOLOGY: unknown
; MOLECULE TYPE: protein
;
US-08-457-459-25

Query Match      13.2% ; Score 50 ; DB 1 ; Length 73 ;
Best Local Similarity 23.9% ; Pred. No. 66 ;
Matches 16 ; Conservative 7 ; Mismatches 10 ; Indels 34 ; Gaps 1 ;

CY      6 DLRKTRIQXKOS-----VQAGTSYRNVLCQ 31
      |||:::|
Db      2 DPKTRIQEVLQGHRLPFTYLVVQVRGAHDQETTHCQVSGLSPEVVGTSRRKAQCA 61
      |||:::|

      32 AIQKSLK 38
      |::|
Db      62 AAEQALK 68

```

```

Db          62 AAEQALK 68

RESULT 6
US-08-555-678-25
; Sequence 25, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,459
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49DUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ. ID NO.: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-555-678-25

Query Match      13.2%; Score 50; DB 1; Length 73;
Best Local Similarity 23.9%; Pred. No. 66;
Matches 16; Conservative 7; Mismatches 10; Indels 34; Gaps 14;

QY      6 DLTKRQKKQS-----VQAGTGYNNVQA 31
      | | | | | | | | | | | | | | | | | |
DB      2 DPTKRLQELQGRHLPPTYLVVQVGEADQEFTHCGVSGISPEVVGTSRRKAQA 61
QY      32 AIQKSLK 38
      | | | | |
DB      62 AAEQALK 68

RESULT 7
PCT-US95-02275-25
; Sequence 25, Application PC/TUS9502275
; GENERAL INFORMATION:
; APPLICANT: Wistar Institute of Anatomy & Biology
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof

```

```

;
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02275
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST99BPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US95-02275-25

Query Match 13.2%; Score 50; DB 5; Length 73;
Best Local Similarity 23.9%; Pred. No. 66;
Matches 16; Conservative 7; Mismatches 10; Indels 34; Gaps 1;

Cy 6 DLKTKIOKYKOS-----VQAGTSYRNVIA 31
Db 2 DPKTRDQVYLGSHLPLTYLVYVVRGAMHDEFTIHCVSGLSPVVGTSRRKAEQA 61
Cy 32 AIQSLK 38
Db 62 AAEQALK 68

RESULT 8
US-09-020-116-5
; Sequence 5, Application US/09020116
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: SUITE 1200, 127 PEACHTREE STREET
; CITY: ATLANTA
; STATE: GA
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```

;
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,116
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLER, MARY L
; REGISTRATION NUMBER: 39,303
; REFERENCE/DOCKET NUMBER: 14014,0285
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-020-116-5

Query Match 13.1%; Score 49.5; DB 3; Length 40;
Best Local Similarity 27.8%; Pred. No. 36;
Matches 10; Conservative 11; Mismatches 6; Indels 9; Gaps 1;

Cy 37 LKDPNNPREPVPKKSIOESAFLOSPERRYMK 72
Db 7 VRDPTSNKQQRFPVS-----QLPGRFOTK 33

RESULT 9
US-09-608-902-5
; Sequence 5, Application US/09608902
; GENERAL INFORMATION:
; APPLICANT: Vonakis, Becky
; APPLICANT: Metzger, Henry
; APPLICANT: Chen, Huaxian
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF
; FILE REFERENCE: 14014,028502
; CURRENT APPLICATION NUMBER: US/09/608,902
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/020,116
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: No. 6353097e = synthetic c
; US-09-608-902-5

Query Match 13.1%; Score 49.5; DB 4; Length 40;
Best Local Similarity 27.8%; Pred. No. 36;
Matches 10; Conservative 11; Mismatches 6; Indels 9; Gaps 1;

Cy 37 LKDPNNPREPVPKKSIOESAFLOSPERRYMK 72
Db 7 VRDPTSNKQQRFPVS-----QLPGRFOTK 33

RESULT 10
US-09-845-917A-39
; Sequence 39, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:-
```

```

; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-39

Query Match          12.9%; Score 49; DB 4; Length 44;
Best Local Similarity 38.1%; Pred. No. 47;
Matches      8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY      26 RNVLQAAIQKSLKDPNNFR 46
DB      18 RHLKSKSLSKIRDISNDFRD 38

RESULT 11
US-09-732-210-1542
; Sequence 1542, Application US/09732210
; Patent No. 6573351
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Miteanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
; SEQ ID NO 1542
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-732-210-1542

Query Match          12.9%; Score 49; DB 4; Length 56;
Best Local Similarity 28.8%; Pred. No. 63;
Matches      15; Conservative 12; Mismatches 13; Indels 12; Gaps 2;

QY      7 LKTKIOKXKQSGAGTSRYNVLQAAIQKSLKDPNNFRPVPKKSIOESPA 58
DB      11 LEPALRRFRKRSVSKTGT-----LQARK--REFYKPSYKAKKSSPA 50

RESULT 12
US-09-045-631-12
; Sequence 12, Application US/09045631B
; Patent No. 6077682
; GENERAL INFORMATION:
; APPLICANT: Inouye, Masayori
; APPLICANT: Park, Heiyous
; APPLICANT: Ikura, Mitsuo
; TITLE OF INVENTION: METHODS OF IDENTIFYING INHIBITORS OF PROTEIN HISTIDINE
; TITLE OF INVENTION: KINASES THROUGH RATIONAL DRUG DESIGN
; FILE REFERENCE: 601-1-082
; CURRENT APPLICATION NUMBER: US/09/045,631B
; CURRENT FILING DATE: 1998-03-19
```

```

; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-045-631-12

Query Match          12.5%; Score 47.5; DB 3; Length 67;
Best Local Similarity 25.0%; Pred. No. 1.2e+02;
Matches      13; Conservative 6; Mismatches 20; Indels 13; Gaps 1;

QY      26 RNVLQAAIQKSLKDPNNFR-----EEPVKKSIOESPAFLPQSI 64
DB      12 RFLMAGVSHDLRPTLRIRLATEMMSEQDGYLAESINKDIECNALIEQFI 63

RESULT 13
US-09-158-843A-12
; Sequence 12, Application US/09158843A
; Patent No. 6162627
; GENERAL INFORMATION:
; APPLICANT: Inouye, Masayori
; APPLICANT: Park, Heiyous
; APPLICANT: Ikura, Mitsuo
; TITLE OF INVENTION: METHOD OF IDENTIFYING INHIBITORS OF SENSOR HISTIDINE
; TITLE OF INVENTION: KINASE THROUGH RATIONAL DRUG DESIGN
; FILE REFERENCE: 601-1-082CIP
; CURRENT APPLICATION NUMBER: US/09/158,843A
; CURRENT FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/078,769
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-158-843A-12

Query Match          12.5%; Score 47.5; DB 3; Length 67;
Best Local Similarity 25.0%; Pred. No. 1.2e+02;
Matches      13; Conservative 6; Mismatches 20; Indels 13; Gaps 1;

QY      26 RNVLQAAIQKSLKDPNNFR-----EEPVKKSIOESPAFLPQSI 64
DB      12 RFLMAGVSHDLRPTLRIRLATEMMSEQDGYLAESINKDIECNALIEQFI 63

RESULT 14
US-09-187-789-62
; Sequence 62, Application US/09187789
; Patent No. 6340740
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: 480140.434CI
; CURRENT APPLICATION NUMBER: US/09/187,789
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-187-789-62

Query Match          11.9%; Score 45; DB 4; Length 31;
Best Local Similarity 58.8%; Pred. No. 92;
Matches      10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

Thu Sep 9 07:46:03 2004

us-10-713-208-6\_copy\_62\_136.rai

Page 6

QY 55 ESEAFLPQSIPEERYKM 71  
|||:|||||  
Db 2 ESEMSDPQPLQEEERYDM 18

RESULT 15  
US-09-139-600-57  
; Sequence 57, Application US/09139600  
; Patent No. 6432628  
; GENERAL INFORMATION:  
; APPLICANT: Alnemri, Emad S.  
; APPLICANT: Fernandez-Alnemri, Teresa  
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
; TITLE OF INVENTION: AND METHOD OF USE  
; FILE REFERENCE: 480140.434  
; CURRENT APPLICATION NUMBER: US/09/139,600  
; CURRENT FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: Patentl Ver. 2.0  
; SEQ ID NO 57  
; LENGTH: 31  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-139-600-57

Query Match 11.9%; Score 45; DB 4; Length 31;  
Best local Similarity 58.8%; Pred. No. 92;  
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
QY 55 ESEAFLPQSIPEERYKM 71  
|||:|||||  
Db 2 ESEMSDPQPLQEEERYDM 18

Search completed: September 9, 2004, 06:18:58  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:19:02 ; Search time 15 Seconds

(without alignments)  
64.128 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193

Perfect score: 62

Sequence: 1 CMPEHRDYDS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 1101

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	21	33.9	9	2 S30494	cat gene leader pe
2	21	33.9	9	2 B24362	chloramphenicol O-
3	20	32.3	10	2 S62208	polyferredoxin - M
4	19	30.6	8	2 S69165	ferredoxin a2 - Ja
5	18	29.0	10	2 F41839	ribosomal protein
6	18	29.0	10	2 A49187	gonadotropin-relea
7	18	29.0	10	2 PC4374	telomeric and tetr
8	17	27.4	10	2 B23565	R-phycocyanin al
9	17	27.4	8	2 S11078	glucose-6-phosphat
10	17	27.4	8	2 I57532	gene Tn10low prote
11	17	27.4	9	2 PT0270	Ig heavy chain CRD
12	17	27.4	10	2 PT0243	Ig heavy chain CRD
13	17	27.4	10	2 PT0291	Ig heavy chain CRD
14	16	25.8	9	2 S36850	Ig heavy chain V r
15	15	24.2	6	2 B44510	hypothetical prote
16	15	24.2	7	2 A15398	choline oxidase (E
17	15	24.2	9	2 T31612	hypothetical prote
18	15	24.2	9	2 A44787	calitWRamide 10
19	15	24.2	9	2 PT0299	Ig heavy chain CRD
20	15	24.2	10	2 PH0944	T-cell receptor be
21	15	24.2	10	2 T14212	cytochrome-c oxida
22	14	22.6	5	2 PT0308	Ig heavy chain CRD
23	14	22.6	6	2 S11556	hydrogenulfite re
24	14	22.6	7	2 B34818	vicilin 57K chain
25	14	22.6	7	2 B45648	Na+-transporting A
26	14	22.6	8	2 A31570	angiotensin-conver
27	14	22.6	8	2 PT0030	inlinase (EC 3.2.
28	14	22.6	9	2 JN0026	sperm-activating p
29	14	22.6	9	2 S13889	phosphoenolpyruvat

30	14	22.6	9	2 JN0027	[Phe-6]-mosact - S
31	14	22.6	9	2 S77984	cytochrome-c oxida
32	14	22.6	10	1 RHPG6	gonadolibertin - pi
33	14	22.6	10	1 RSHSG	gonadolibertin - sh
34	14	22.6	10	1 RHA01	gonadolibertin I -
35	14	22.6	10	2 A61337	caerulein - frog (
36	14	22.6	10	2 S74176	glucokinase (EC
37	14	22.6	10	2 UN0025	mosact - sea urchi
38	14	22.6	10	2 A21114	gonadolibertin - ch
39	14	22.6	10	2 PH0948	T-cell receptor be
40	14	22.6	10	2 I52645	gene B-50 protein
41	13	21.0	4	2 I54357	schwannomin - mous
42	13	21.0	5	2 S68326	blood cell protein
43	13	21.0	6	2 B56979	collagen alpha 1(I
44	13	21.0	8	2 P00012	cholecystokinin -
45	13	21.0	8	2 A43001	cholecystokinin -

#### ALIGNMENTS

##### RESULT 1

S30494 cat gene leader peptide - Streptococcus agalactiae plasmid pIP501

C/Species: Streptococcus agalactiae

C/Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 24-Sep-1999

C/Accession: S30494

R/Trieu-Cloc, P.; de Cespedes, G.; Horaud, T.

A/Title: Nucleotide sequence of the chloramphenicol resistance determinant of the Strep.

A/Reference number: J01950; MUID:93096867; PMID:1461942

A/Accession: S30494

A/Molecule type: DNA

A/Residues: 1-9 <TRI>

A/Cross-references: EMBL:X65462; NID:949071; PIDN:CAA46454.1; PID:9581554

C/Genetics:

A/Genome: plasmid pIP501

C/superfamily: unassigned leader peptides

Query Match 33.9% Score 21; DB 2; Length 9;

Best Local Similarity 44.4% Pred. No. 2.8e+05; Indels 0; Gaps 0;

Matches 4; Conservative 1; Mismatches 4;

QY 2 MPEHRDYDS 10

DB 1 MKKSEDDYS 9

##### RESULT 2

B24362 chloramphenicol O-acetyltransferase leader peptide - Staphylococcus aureus plasmid pUB1

C/Species: Staphylococcus aureus

C/Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 11-May-2000

C/Accession: B24362

R/Buckner, R.; Matczura, H.

A/Title: Regulation of the inducible chloramphenicol acetyltransferase gene of the Stap

A/Reference number: A24362; MUID:86081739; PMID:3865770

A/Accession: B24362

A/Molecule type: DNA

A/Residues: 1-9 <BRU>

A/Cross-references: GB:X02872; NID:946536; PIDN:CAA26630.1; PID:9581555

C/Comment: Ribosome stalling in the translation of this leader peptide, caused by the e

nsation of the chloramphenicol O-acetyltransferase from a ribosome binding site located

A/Genetics:

A/Genome: plasmid

C/superfamily: unassigned leader peptides

Query Match 33.9% Score 21; DB 2; Length 9;

Best Local Similarity 44.4% Pred. No. 2.8e+05; Indels 0; Gaps 0;

Matches 4; Conservative 1; Mismatches 4;

QY 2 MPEHRDYDS 10

Db 1 MKKSEDDYSS 9

## RESULT 3

polyferredoxin - Methanosarcina barkeri (fragment)  
 C/Species: Methanosarcina barkeri  
 C/Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Mar-1997  
 C/Accession: S62208  
 R/Vorholt, J.A.; Vaupel, M.; Thauer, R.K.  
 Eur. J. Biochem. 236: 309-317, 1996  
 A/Title: A polyferredoxin with eight [4Fe-4S] clusters as a subunit of molybdenum formyl  
 A/Reference number: S62194; MUID:96184912; PMID:8617280  
 A/Accession: S62208  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-10 <VOR>

Query Match 32.3%; Score 20; DB 2; Length 10;  
 Best Local Similarity 75.0%; Pred. No. 1e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYDS 10  
 |||:  
 4 DYDN 7

## RESULT 4

ferredoxin a2 - Japanese radish (fragment)  
 C/Species: Kaiware daikon (Japanese radish)  
 C/Date: 10-Mar-1998 #sequence\_revision 10-Mar-1998 #text\_change 17-Apr-1998  
 C/Accession: S69165  
 R/Odeta, S.; Nishimura, M.; Nagai, K.; Sakihama, N.; Shin, M.  
 Arch. Biochem. Biophys. 316: 797-802, 1995  
 A/Title: Four ferredoxins from Japanese radish leaves.  
 A/Reference number: S69164; MUID:9516867; PMID:7864635  
 A/Accession: S69165  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-8 <OBA>  
 C/Keywords: zfe-25; electron transfer; iron-sulfur protein

Query Match 30.6%; Score 19; DB 2; Length 8;  
 Best Local Similarity 60.0%; Pred. No. 2.8e+05;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 HRDYD 9  
 |||:  
 Db 2 HRED 6

## RESULT 5

F41839  
 ribosomal protein L16 - Acholeplasma laidlawii (fragment)  
 C/Species: Acholeplasma laidlawii  
 C/Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 18-Nov-1994  
 C/Accession: F41839  
 R/Lim, P.O.; Sears, B.B.  
 J. Bacteriol. 174: 2606-2611, 1992  
 A/Title: Evolutionary relationships of a plant-pathogenic mycoplasma-like organism and Ac  
 A/Reference number: A41839; MUID:92210505; PMID:1556079  
 A/Accession: F41839  
 A/Status: preliminary; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-10 <LIM>  
 A/Cross-references: GB:W74471  
 C/Genetics:  
 A:Gene: rpl16  
 C/Keywords: protein biosynthesis; ribosome

Query Match 29.0%; Score 18; DB 2; Length 10;  
 Best Local Similarity 42.9%; Pred. No. 2.5e+03;

Matches 3; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 MPEHRYD 8  
 |||:  
 Db 3 MPKRTKY 9

## RESULT 6

gonadotropin-releasing hormone III - sea lamprey  
 C/Species: Petromyzon marinus (sea lamprey)  
 C/Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-Mar-1995  
 C/Accession: A49187  
 R/Sower, S.A.; Chiang, Y.C.; Lovaas, S.; Conlon, J.M.  
 Endocrinology 132, 1125-1131, 1993  
 A/Title: Primary structure and biological activity of a third gonadotropin-releasing  
 A/Reference number: A49187; MUID:93178316; PMID:8440174  
 A/Accession: A49187  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-10 <SOW>  
 A/Experimental source: brain  
 A/Note: sequence extracted from NCBI backbone (NCBIP:126381)

Query Match 29.0%; Score 18; DB 2; Length 10;  
 Best Local Similarity 50.0%; Pred. No. 2.5e+03;  
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 EHRDYD 9  
 |||:  
 Db 1 EHWSD 6

## RESULT 7

telomeric and tetraplex DNA binding protein qTBP42 IV - rat (fragment)  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 28-Oct-1997 #sequence\_revision 28-Oct-1997 #text\_change 07-May-1999  
 C/Accession: PC4374  
 R/Saris, G.; Weisman-Shomer, P.; Fry, M.  
 Biochem. Biophys. Res. Commun. 237: 617-623, 1997  
 A/Title: Telomeric and tetraplex DNA binding properties of qTBP42: A homologue of the  
 A/Reference number: PC4371; MUID:97445086; PMID:9299414  
 A/Accession: PC4374  
 A/Molecule type: protein  
 A/Residues: 1-10 <SAR>  
 C/Comment: This protein binds either strand of the telomeric DNA as well as unimolecu

Query Match 29.0%; Score 18; DB 2; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+03;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EHR 6  
 |||:  
 Db 7 EHR 9

## RESULT 8

B22565  
 R-phycoerythrin alpha-2 chain - red alga (Gastrocloonium coulteri) (fragment)  
 C/Species: Gastrocloonium coulteri  
 C/Date: 07-Mar-1988 #sequence\_revision 07-Mar-1988 #text\_change 23-Mar-1993  
 C/Accession: B22565  
 R/Klotz, A.V.; Glazer, A.N.  
 J. Biol. Chem. 260: 4856-4863, 1985  
 A/Title: Characterization of the bilin attachment sites in R-phycoerythrin.  
 A/Reference number: A22565; MUID:85182601; PMID:3886644  
 A/Accession: B22565  
 A/Molecule type: protein  
 A/Residues: 1-5 <KLO>

Query Match 27.4%; Score 17; DB 2; Length 5;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+05;



Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CMP 3  
|||  
Db 2 CVP 4

## RESULT 9

S11078  
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) - yeast (*Pichia jadinii*) (fragment)  
C/Species: *Pichia jadinii*, *Candida utilis*  
C/Date: 30-Jun-1991 #sequence\_revision 30-Sep-1991 #text\_change 05-Aug-1994  
C/Accession: S11078  
R/Egestad, B.; Estoniun, M.; Danielsson, O.; Persson, B.; Cedertund, E.; Kaiser, R.; Hol  
FBS Lett. 269, 194-196, 1990  
A/Title: Fast atom bombardment mass spectrometry and chemical analysis in determinations  
A/Reference number: S11074; MUID:90355571; PMID:2387402  
A/Accession: S11078  
A/Molecule type: proteoin  
A/Residues: 1-8 <EGS>  
A/Note: the source is designated as *Pichia jadinii*  
C/Keywords: acetylated amino end; oxidoreductase; pentose phosphate pathway  
F:/Modified site: acetylated amino end (Ser) #status experimental

Query Match 27.4%; Score 17; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YDS 10  
|||  
Db 2 YDS 4

## RESULT 10

S15732  
gene Tnslow protein - rat (fragment)  
C/Species: *Rattus sp.* (rat)  
C/Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 28-Feb-1997  
C/Accession: S15732  
R/Banerjee-Basu, S.; Buonanno, A.  
Mol. Cell. Biol. 13, 7019-7028, 1993  
A/Title: cis-acting sequences of the rat tropomyosin I slow gene confer tissue- and develop  
A/Reference number: S15732; MUID:94019373; PMID:8411291  
A/Accession: S15732  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-8 <RBS>  
A/Cross-references: GB:S66172; NID:9432603  
C/Genetics:  
A/Gene: Tnslow

Query Match 27.4%; Score 17; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MPE 4  
|||  
Db 1 MPE 3

## RESULT 11

PT0270  
Ig heavy chain CRD3 region (clone 3-100) - human (fragment)  
C/Species: *Homo sapiens* (man)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C/Accession: PT0270  
R/Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A/Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A/Reference number: PT0222; MUID:91108337; PMID:1899102  
A/Accession: PT0270  
A/Molecule type: DNA  
A/Residues: 1-9 <YAM>

A/Experimental source: B lymphocyte  
C/Keywords: heterotetramer; immunoglobulin

Query Match 27.4%; Score 17; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YDS 10  
|||  
Db 4 YDS 6

## RESULT 12

PT0243  
Ig heavy chain CRD3 region (clone 2-103A) - human (fragment)  
C/Species: *Homo sapiens* (man)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C/Accession: PT0243  
R/Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A/Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A/Reference number: PT0222; MUID:91108337; PMID:1899102  
A/Accession: PT0243  
A/Molecule type: DNA  
A/Residues: 1-10 <YAM>  
A/Experimental source: B lymphocyte  
C/Keywords: heterotetramer; immunoglobulin

Query Match 27.4%; Score 17; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YDS 10  
|||  
Db 4 YDS 6

## RESULT 13

PT0291  
Ig heavy chain CRD3 region (clone 4-115B) - human (fragment)  
C/Species: *Homo sapiens* (man)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Aug-1996  
C/Accession: PT0291  
R/Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.  
J. Exp. Med. 173, 395-407, 1991  
A/Title: Preferential utilization of specific immunoglobulin heavy chain diversity and  
A/Reference number: PT0222; MUID:91108337; PMID:1899102  
A/Accession: PT0291  
A/Molecule type: DNA  
A/Residues: 1-10 <YAM>  
A/Experimental source: B lymphocyte  
C/Keywords: heterotetramer; immunoglobulin

Query Match 27.4%; Score 17; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 3.8e+03;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YDS 10  
|||  
Db 8 YDS 10

## RESULT 14

S36850  
Ig heavy chain V region - mouse  
C/Species: *Mus musculus* (house mouse)  
C/Date: 06-Jan-1994 #sequence\_revision 01-Dec-1995 #text\_change 05-Nov-1999  
C/Accession: S36850  
R/Jacob, J.; Keisler, G.  
submitted to the EMBL Data Library, July 1992  
A/Description: In situ studies on the primary immune response to (4-hydroxy-3-nitrophen  
A/Reference number: S25024  
A/Accession: S36850

A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-9 <UAC>  
A;Cross-references: EMBL:X67387; NID:G50113; PIDN:CAA47799.1; PID:es1594; PID:G1333871  
C;Keywords: heterotetramer; immunoglobulin

Query Match 25.8%; Score 16; DB 2; Length 9;  
Best Local Similarity 75.0%; Pred. No. 2.8e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 7 DYDS 10  
|||  
Db 4 DYGS 7

RESULT 15  
B44510  
hypothetical protein C (mler 3' region) - Lactococcus lactis (fragment)  
C;Species: Lactococcus lactis  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 30-Sep-1993  
C;Accession: B44510  
R;Renault, P.; Galliardin, C.; Heesloc, H.  
J. Bacteriol. 171, 3108-3114, 1989  
A;Title: Product of the Lactococcus lactis gene required for malolactic fermentation is  
A;Reference number: A44510; XUID:89255069; PMID:2498286  
A;Accession: B44510  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-6 <REN>  
A;Cross-references: EMBL:M90762

Query Match 24.2%; Score 15; DB 2; Length 6;  
Best Local Similarity 66.7%; Pred. No. 2.8e+05;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 RDY 8  
:|  
Db 3 KDY 5

Search completed: September 9, 2004, 06:22:10  
Job time : 16 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:18:27 ; Search time 10 seconds  
(without alignments)  
52.070 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193  
Perfect score: 62  
Sequence: 1 CMPEHRDYDS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 371

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	33.9	9	1 LPCA_STAUV	P36884 Staphylococ
2	18	29.0	10	1 RL16_ACHLA	P29221 achyloplasm
3	17	27.4	8	1 AL15_CYDPO	P82156 Cydia pomon
4	17	27.4	10	1 UN05_RAT	P65573 Rattus norv
5	15	24.2	7	1 CHOX_ALGSP	P16101 alcaligenes
6	15	24.2	9	1 FARA_CALVO	P41865 Calliphora
7	15	24.2	10	1 GON3_PETVA	P30948 petromyzon
8	14	22.6	8	1 ACT_THUAL	P18691 thunnus alb
9	14	22.6	9	1 COX6_MAIZE	P80632 zea mays (m
10	14	22.6	9	1 COXE_THUOB	P80975 thunnus obe
11	14	22.6	9	1 PAR3_MACRS	P83276 macrobachi
12	14	22.6	9	1 PAR5_PENNO	P83320 penaeus mon
13	14	22.6	9	1 MOSF_CLYVA	P19853 clypeaster
14	14	22.6	9	1 MOSH_CLYVA	P19852 clypeaster
15	14	22.6	10	1 CA12_LITCI	P82065 litorea cit
16	14	22.6	10	1 CAER_LITXA	P66264 litorea xan
17	14	22.6	10	1 GON1_AL1MT	P37041 alligator m
18	14	22.6	10	1 GON3_ONCKE	P20367 oncothychnu
19	14	22.6	10	1 MOSQ_CLYVA	P19962 clypeaster
20	13	21.0	6	1 ASP2_LACSN	P82168 lactobacill
21	13	21.0	7	1 AL17_CYDPO	P82158 cydia pomon
22	13	21.0	7	1 TY51_LITRU	P82065 litorea rub
23	13	21.0	8	1 AL15_CALVO	P41841 calliphora
24	13	21.0	8	1 CCKN_MACRU	P30369 macropus eu
25	13	21.0	8	1 COXG_RAT	P80430 rattus norv
26	13	21.0	8	1 GURR_HUMAN	P02729 homo sapien
27	13	21.0	9	1 BS43_SERPL	P83375 serrata pi
28	13	21.0	9	1 NSK1_SARBU	P41492 sarcophaga
29	13	21.0	9	1 UPAT_HUMAN	P30093 homo sapien
30	13	21.0	10	1 AKHX_LOCOM	P81026 locusta mig
31	13	21.0	10	1 CATB_SHEEP	P83205 ovis aries
32	13	21.0	10	1 EST1_SCHRA	P81012 schizaphis
33	13	21.0	10	1 GON1_CHEBR	P80677 chelyosoma

34	13	21.0	10	1 GS09_BACSU	P80243 bacillus su
35	13	21.0	10	1 LSK2_LEIMA	P09039 leucophaea
36	13	21.0	10	1 FORB_MERTM	P80901 methanobact
37	13	21.0	10	1 TWOF_AEDAE	P19425 aedes aegypt
38	12	19.4	6	1 OVM_LEPDE	P42965 lepidoptera
39	12	19.4	7	1 CARP_MYTED	P10420 mytilus edu
40	12	19.4	7	1 E105_LITRU	P82101 litorea rub
41	12	19.4	8	1 UN09_RAT	P56575 rattus norv
42	12	19.4	9	1 FARA_MACRS	P83281 macrobachi
43	12	19.4	9	1 SNP_STOVA	P24047 stomopneute
44	12	19.4	9	1 UHA2_HUMAN	P40929 homo sapien
45	12	19.4	10	1 AN3_PRUSE	P29261 prunus sero

## ALIGNMENTS

```

RESULT 1
ID      LPCA_STAUV      STANDARD;      PRT;      9 AA.
AC      P36884;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-OCT-1994 (Rel. 30, Last annotation update)
DE      Chloramphenicol resistance leader peptide.
OS      Staphylococcus aureus, and
OG      Staphylococcus agalactiae.
OC      Plasmid pSCS6, Plasmid pSCS7, Plasmid pUBA12, and Plasmid pIP501.
OX      Bacteria; Firmicutes; Bacillales; Staphylococcus.
[1]
RN      NCBI_TaxId=1280, 1311;
RP      SEQUENCE FROM N.A.
RC      SPECIES=S. aureus; STRAIN=436; PLASMID=pSCS7;
RX      MEDLINE=92207652; PubMed=1929326;
RA      Schwarz S.; Cardoso M.;
RT      "Nucleotide sequence and phylogeny of a chloramphenicol
RT      acetyltransferase encoded by the plasmid pSCS7 from Staphylococcus
RT      aureus."
RL      Antimicrob. Agents Chemother. 35:1551-1556(1991).
[2]
RN      SEQUENCE FROM N.A.
RP      SPECIES=S. aureus; PLASMID=pSCS6;
RX      MEDLINE=92288047; PubMed=1517170;
RA      Cardoso M.; Schwarz S.;
RT      "Nucleotide sequence and structural relationships of a
RT      chloramphenicol acetyltransferase encoded by the plasmid pSCS6 from
RT      Staphylococcus aureus."
RL      J. Appl. Bacteriol. 72:289-293(1992).
[3]
RN      SEQUENCE FROM N.A.
RP      SPECIES=S. aureus; PLASMID=pUBA12;
RX      MEDLINE=86081739; PubMed=3865770;
RA      Brueckner R.; Matzura H.;
RT      "Regulation of the inducible chloramphenicol acetyltransferase gene
RT      of the Staphylococcus aureus plasmid pUBA12."
RL      EMBO J. 4:2295-2300(1985).
[4]
RN      SEQUENCE FROM N.A.
RP      SPECIES=S. agalactiae; PLASMID=pIP501;
RX      MEDLINE=93096867; PubMed=1461942;
RA      Trieu-Quoc P.; de Cespedes G.; Horaud T.;
RT      "Nucleotide sequence of the chloramphenicol resistance determinant of
RT      the streptococcal plasmid pIP501."
RL      Plasmid 28:272-276(1992).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

DR EMBL; M58515; AAA26612.1; -  
 DR EMBL; M58516; AAA16528.1; -  
 DR EMBL; X02872; CAA26630.1; -  
 DR EMBL; X60827; CAA43217.1; -  
 DR EMBL; X65462; CAA46454.1; -  
 DR PIR; B24362; B24362.  
 DR PIR; S30494; S30494.  
 KM Leader peptide; Antibiotic resistance; Plasmid.  
 SQ SEQUENCE 9 AA; 1074 MW; 5D9CAB5AA505333 CRC64;

Query Match 33.9%; Score 21; DB 1; Length 9;  
 Best Local Similarity 44.4%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 MPEHRDYS 10  
 |||  
 DB 1 MKKSEDTSS 9

## RESULT 2

RL16\_ACHUA STANDARD; PRT; 10 AA.  
 AC P29221;  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE 50S ribosomal protein L16 (Fragment).  
 GN RPLP.  
 OS Acholeplasma laidlawii.  
 CC Bacteria; Firmicutes; Mollicutes; Acholeplasmatales;  
 OC Acholeplasmataceae; Acholeplasma.  
 OX NCBI\_TaxID=2148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92210505; Pubmed=1556079;  
 RA Lin P.O., Sears B.B.;  
 RT "Evolutionary relationships of a plant-pathogenic mycoplasma-like  
 organism and Acholeplasma laidlawii deduced from two ribosomal protein  
 gene sequences."  
 RL J. Bacteriol. 174:2606-2611(1992).  
 CC -1- FUNCTION: This protein binds directly to 23S ribosomal RNA and is  
 located at the site of the peptidyltransferase center (By  
 similarity).  
 CC -1- SIMILARITY: Belongs to the L16P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M74771; AAA21914.1; -  
 DR PIR; F41839; F41839.  
 DR InterPro: IPR000114; Ribosomal L16.  
 DR PROSITE; PS00586; RIBOSOMAL\_L16\_1; PARTIAL.  
 DR PROSITE; PS00701; RIBOSOMAL\_L16\_2; PARTIAL.  
 KM Ribosomal protein; rRNA-binding.  
 FT NON TER 10  
 SQ SEQUENCE 10 AA; 1324 MW; B3386A21B4032766 CRC64;

Query Match 29.0%; Score 18; DB 1; Length 10;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+03;  
 Matches 3; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 MPEHRDY 8  
 |||  
 DB 3 MKKRTKY 9

## RESULT 3

ALL5\_CYPDO

ID ALL5\_CYPDO STANDARD; PRT; 8 AA.  
 AC P82156;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Cydiastatin 5.  
 OS Cydia pomonella (Codling moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Tortricidae; Tortricidae; Olethreutinae; Cydia.  
 OX NCBI\_TaxID=82600;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Larva;  
 RX MEDLINE=96054539; Pubmed=9392829;  
 RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,  
 RA Davey M., East P.D., Inorpe A.;  
 RT "Lepidopteran peptides of the allatostatin superfamily."  
 RL Peptides 18:1301-1309(1997).  
 CC -1- SIMILARITY: Belongs to the allatostatin family.  
 KM Neuropeptide; Amidation.  
 FT MOD RES 8  
 SQ SEQUENCE 8 AA; 898 MW; 922879CABB5640D CRC64;

Query Match 27.4%; Score 17; DB 1; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 RPYD 9  
 |||  
 DB 2 RGYD 5

RESULT 4  
 IDH05\_RAT STANDARD; PRT; 10 AA.  
 AC P56573;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Unknown protein from 2D-page of heart tissue (Spot P5) (Fragment).  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=Mistar; TISSUE=Heart;  
 RA Li X.-P., Pleisner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,  
 RA Umplupf P.R.;  
 RL Submitted (SEP-1998) to Swiss-Prot.  
 CC -1- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown  
 CC protein is: 8.3, its MW is: 30 KDa.  
 FT UNSURE 9  
 FT NON TER 10  
 SQ SEQUENCE 10 AA; 1269 MW; 00CBA4BB46CSBAB CRC64;

Query Match 27.4%; Score 17; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YDS 10  
 |||  
 DB 3 YDS 5

RESULT 5  
 ID CHO\_X\_ALCSP STANDARD; PRT; 7 AA.  
 AC P16101;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last annotation update)

```

DE Choline oxidase (EC 1.1.3.17) (Fragment).
OS Alcaligenes sp.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OX NCBI_TaxID=512;
RN [1]
RP SEQUENCE.
RX MEDLINE=81006769; PubMed=6997283;
RA Ohta-Fukuyama M., Miyake Y., Emi S., Yamano T.;
RT "Identification and properties of the prosthetic group of choline
RT oxidase from Alcaligenes sp."
RC J. Biochem. 88:197-203(1980).
DR -I- CATALYTIC ACTIVITY: Choline + O(2) = betaine aldehyde + H(2)O(2).
KW Oxidoreductase.
SQ SEQUENCE 7 AA; 839 MW; 7415B1E457644AC0 CRC64;

Query Match
Best Local Similarity 24.2%; Score 15; DB 1; Length 7;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PER 5
DB 3 PNH 5

RESULT 6
FARA_CALVO
ID FARA_CALVO STANDARD; PRT; 9 AA.
AC P1865;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Calliphoridae; Calliphora.
OS Calliphora vomitoria (Blue blowfly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Phrygora;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
OC Calliphoridae; Calliphora.
OX NCBI_TaxID=27454;
RN [1]
RP SEQUENCE.
RX TISSUE=Thoracic ganglion;
MEDLINE=92196111; PubMed=1549595;
RA Duvé H., Johnsen A.H., Sewell J.C., Scott A.G., Orchard I.,
RA Rehfeld J.F., Thorpe A.;
RT "Isolation, structure, and activity of -Phe-Met-Arg-Phe-NH2
RT neuropeptides (designated calliphorin) from the blowfly
RT Calliphora vomitoria."
RL Proc. Natl. Acad. Sci. U.S.A. 89:2326-2330(1992).
CC -I- SIMILARITY: Belongs to the PARP (FMRamide related peptide)
CC family.
DR PIR: A44787; A44787.
KW Neuropeptide; Amidation.
FT MOD_RES 9 AMIDATION.
FT UNSURE 1 OR S OR A.
SQ SEQUENCE 9 AA; 1183 MW; 29D0699CA840457 CRC64;

Query Match
Best Local Similarity 24.2%; Score 15; DB 1; Length 9;
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 HRDY 8
DB 3 NRDF 6

RESULT 7
GON3_PETMA
ID GON3_PETMA STANDARD; PRT; 10 AA.
AC P30948;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gonadoliberin III (Gonadotropin-releasing hormone III) (GNRH-III)
DE (Luliberin III).
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE.
RX MEDLINE=93178316; PubMed=8440174;
RA Sower S.A., Chiang Y.-C., Lovas S., Conlon J.M.;
RT "Primary structure and biological activity of a third gonadotropin-
RT releasing hormone from lamprey brain."
RL Endocrinology 132:1125-1131(1993).
CC -I- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates
CC the secretion of both luteinizing and follicle-stimulating
CC hormones.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: Belongs to the GnRH family.
DR InterPro: IPR02012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Hormone; Amidation; Hypothalamus; Pyrrolidone carboxylic acid.
FT MOD_RES 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 AMIDATION.
SQ SEQUENCE 10 AA; 1277 MW; 284B36237A1F5A3 CRC64;

Query Match
Best Local Similarity 24.2%; Score 15; DB 1; Length 10;
Matches 2; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 EHRDYD 9
DB 1 QWMSHD 6

RESULT 8
ACT_THUAL
ID ACT_THUAL STANDARD; PRT; 8 AA.
AC P18691;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Arginensin-converting enzyme inhibitor.
OS Thunnus albacares (Yellowfin tuna) (Neothunnus macropterus).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
OC Scombridae; Thunnus.
OX NCBI_TaxID=8236;
RN [1]
RP SEQUENCE.
RX TISSUE=Muscle;
MEDLINE=88326322; PubMed=3415688;
RA Kohama Y., Matsunoto S., Oka H., Teramoto T., Okabe M., Minura T.;
RT "Isolation of angiotensin-converting enzyme inhibitor from tuna
RT muscle."
RL Biochem. Biophys. Res. Commun. 155:332-337(1988).
DR PIR: A31570; A31570.
SQ SEQUENCE 8 AA; 953 MW; 6AA863733051F1B7 CRC64;

Query Match
Best Local Similarity 22.6%; Score 14; DB 1; Length 8;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PER 5
DB 1 PTH 3

RESULT 9
UC26_MAIZE

```

ID UC26 MAIZE STANDARD; PRT; 8 AA.  
 AC P80632;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Unknown protein from 2D-page of etiolated coleoptile (Spot 907) (Fragment).  
 OS Zea mays (Maize).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxId=4577;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Coleoptile;  
 RA Touzet P., Ricciardi F., Morin C., Damerval C., Huet J.-C., Pennolet J.-C., Zivy M., de Vienne D.;  
 RT "The maize two dimensional gel protein database: towards an integrated genome analysis program."  
 RL Theor. Appl. Genet. 93:997-1005(1996).  
 CC -1- MISCELLANEOUS: On the 2D-gel the determined pI of this unknown protein is: 7.0, its MW is: 57.2 KDa.  
 DR Maize-2DPAGE; P80632; COLEOPTILE.  
 FT NON\_TER 1 1  
 FT MOD\_RES 8 8  
 SQ SEQUENCE 8 AA; 990 MW; 9639D6DAB417681D CRC64;  
 Query Match 22.6%; Score 14; DB 1; Length 8;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 EPRD 7  
 DB 2 EPRD 5

RESULT 10  
 COXE\_THUOB STANDARD; PRT; 9 AA.  
 ID COXE\_THUOB  
 AC P80975;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE Cytochrome c oxidase polypeptide via (EC 1.9.3.1) (Fragment).  
 OS Thynnus obesus (Bigeye tuna).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; OC Acanthomorpha; Acanthopterygii; Perciformes; Scombroidei; OC Scombridae; Thynnus.  
 OX NCBI\_TaxId=8241;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Heart;  
 RX MEDLINE=97454291; PubMed=9310366;  
 RA Arnold S., Lee I., Kim M., Song E., Linder D., Lottspeich F., Kadenbach B.;  
 RT "The subunit structure of cytochrome c oxidase from tuna heart and liver."  
 RL Eur. J. Biochem. 248:59-103(1997).  
 CC -1- FUNCTION: This protein is one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the terminal oxidase in mitochondrial electron transport.  
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
 CC -1- SIMILARITY: Belongs to the cytochrome c oxidase via family.  
 DR InterPro: IPR001349; COX6A.  
 DR PROSITE: PS01329; COX6A; PARTIAL.  
 KW Oxidoreductase; Inner membrane; Mitochondrion.  
 FT NON\_TER 1 1  
 FT MOD\_RES 9 9  
 SQ SEQUENCE 9 AA; 1136 MW; 62E072C9CB0776DB CRC64;

Query Match 22.6%; Score 14; DB 1; Length 9;  
 Best Local Similarity 50.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 EPRDY 8  
 DB 4 EPRDY 9

RESULT 11  
 FARS\_MACRS STANDARD; PRT; 9 AA.  
 ID FARS\_MACRS  
 AC P83276;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE FMRFamide-like neuropeptide FLP3 (NPKNPLR-amide).  
 OS Macrobrachium rosenbergii (Giant fresh water prawn).  
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea; OC Palaemonoidea; Palaemonidae; Macrobrachium.  
 OX NCBI\_TaxId=79674;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Eyestalk;  
 RA Sithigorngul P., Sarathongkum W., Jaidechoey S., Longyant S., Sithigorngul W.;  
 RT "Novel FMRFamide-like neuropeptides from the eyestalk of the giant freshwater prawn Macrobrachium rosenbergii."  
 RL Comp. Biochem. Physiol. 120B:587-595(1998).  
 CC -1- MASS SPECTROMETRY: MW=1215.4; METHOD=MALDI.  
 CC -1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide) family.  
 DR GO: GO:0007218; P:neuropeptide signaling pathway; TAS.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 9 9  
 SQ SEQUENCE 9 AA; 1216 MW; 18220729C4433AB CRC64;

Query Match 22.6%; Score 14; DB 1; Length 9;  
 Best Local Similarity 66.7%; Pred. No. 1.4e+05;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYD 9  
 DB 1 NYD 3

RESULT 12  
 FARS\_PENMO STANDARD; PRT; 9 AA.  
 ID FARS\_PENMO  
 AC P83320;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE FMRFamide-like neuropeptide FLP5 (SMPSRLR-amide).  
 OS Penaeus monodon (Penaeid shrimp).  
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea; OC Penaeidae; Penaeus.  
 OX NCBI\_TaxId=6687;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE=Eyestalk;  
 RX MEDLINE=21956277; PubMed=11959015;  
 RA Sithigorngul P., Pupunen J., Krungkarn C., Longyant S., Chaitvutthangkura P., Sithigorngul W., Pansom A.;  
 RT "Seven novel FMRFamide-like neuropeptide sequences from the eyestalk of the giant tiger prawn Penaeus monodon."  
 RL Comp. Biochem. Physiol. 131B:325-337(2002).  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MASS SPECTROMETRY: MW=1121.1; METHOD=MALDI.  
 CC -1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)

CC family.  
DR GO; GO:0007218; P:neuropeptide signaling pathway; TAS.  
KW Neuropeptide; Amidation.  
FT MOD RES 9  
SQ SEQUENCE 9 AA; 1106 MW; B60307340735A766 CRC64;

Query Match 22.6%; Score 14; DB 1; Length 9;  
Best Local Similarity 60.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 MEHR 6  
DB 2 MSLR 6

RESULT 13  
MOSH\_CLYYA STANDARD; PRT; 9 AA.  
AC P19853;  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DE [Phe-6]-mosact.  
OS Clypeaster japonicus (Sand dollar).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidae; Euechinoidea; Gnathostomata; Clypeasteroidea;  
OC Clypeasteridae; Clypeaster.  
NCBI\_TaxID=7644;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Egg jelly;  
RA Suzuki N., Kurita M., Yoshino K.I., Kajitara H., Nomura K.,  
Yamaguchi M.;  
RT "Purification and structure of mosact and its derivatives from the  
egg jelly of the sea urchin Clypeaster japonicus.";  
Zool. Sci. 4:649-656(1987).  
CL -1- FUNCTION: Stimulates sperm respiration and motility.  
DR PIR: JN0027; JN0027.  
SQ SEQUENCE 9 AA; 924 MW; 9324579CDBSAB5 CRC64;

Query Match 22.6%; Score 14; DB 1; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 DYDS 10  
DB 1 DSDS 4

RESULT 14  
MOSH\_CLYYA STANDARD; PRT; 9 AA.  
AC P19852;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Sperm-activating peptide SAP-D ([His]-mosact).  
OS Clypeaster japonicus (Sand dollar).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidae; Euechinoidea; Gnathostomata; Clypeasteroidea;  
OC Clypeasteridae; Clypeaster.  
NCBI\_TaxID=7644;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Egg jelly;  
RA Suzuki N., Kurita M., Yoshino K.I., Kajitara H., Nomura K.,  
Yamaguchi M.;  
RT "Purification and structure of mosact and its derivatives from the  
egg jelly of the sea urchin Clypeaster japonicus.";  
Zool. Sci. 4:649-656(1987).  
CL [2]  
RP BROMINATION OF HIS-6.  
RX MEDLINE=91167743; PubMed=2076468;

RA Takao T., Yoshino K., Suzuki N., Shimomishi Y.;  
RT "Analysis of post-translational modifications of proteins by accurate  
RT mass measurement in fast atom bombardment mass spectrometry.";  
RL Biomed. Environ. Mass Spectrom. 19:705-712(1990)  
CC -1- FUNCTION: Stimulates sperm respiration and motility.  
DR PIR: JN0026; JN0026.  
KW Bromination.  
FT MOD RES 6  
SQ SEQUENCE 9 AA; 914 MW; 93245721BDCSAB5 CRC64;

Query Match 22.6%; Score 14; DB 1; Length 9;  
Best Local Similarity 75.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 DYDS 10  
DB 1 DSDS 4

RESULT 15  
CA12\_LITCI STANDARD; PRT; 10 AA.  
AC P82086;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Caerulein 1.2/1.2Y4.  
OS Litoria citropa (Australian blue mountains tree frog), and  
Litoria splendida (Magnificent tree frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylidae;  
OC Pelodyadinae; Litoria.  
NCBI\_TaxID=94770; 30345;

RP SEQUENCE, AND MASS SPECTROMETRY (CAERULEIN 1.2 AND 1.2Y4).  
RC SPECIES=L. citropa; TISSUE=Skin secretion;  
RX MEDLINE=20057701; PubMed=10589099;  
RA Wabnitz P.A., Boyle J.H., Tyler M.J.;  
RT "Caerulein-like peptides from the skin glands of the Australian blue  
mountains tree frog Litoria citropa. Part 1. Sequence determination  
using electrospray mass spectrometry.";  
RL Rapid Commun. Mass Spectrom. 13:2498-2502(1999).  
RN [2]

RP SEQUENCE, AND MASS SPECTROMETRY (CAERULEIN 1.2).  
RC SPECIES=L. splendida; TISSUE=Skin secretion;  
RX MEDLINE=20069371; PubMed=1601876;  
RA Wabnitz P.A., Boyle J.H., Tyler M.J., Wallace J.C., Smith B.P.;  
RT "Differences in the skin peptides of the male and female Australian  
tree frog Litoria splendida. The discovery of the aquatic male sex  
pheromone splendipherin, together with Phe8 caerulein and the  
antibiotic peptide caerin 1.10.";  
RL Eur. J. Biochem. 267:269-275(2000).  
CC -1- FUNCTION: Hypotensive neuropeptide (Probable).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Skin dorsal glands.  
CC -1- PTM: Isoform 1.2Y4 differs from isoform 1.2 in not being  
sulfated.  
CC -1- MASS SPECTROMETRY: MW=1366; METHOD=Electrospray.  
CC -1- SIMILARITY: Belongs to the gastrin/cholecystokinin family.  
DR Interpro, IPR01651; Gastrin.  
DR PROSITE, PS00259; GASTRIN; FALSE\_NEG.  
KW Amphibian defense peptide; Hypotensive agent; Amidation; Sulfation;  
KW Pyrrolidone carboxylic acid.  
FT MOD RES 1 1  
FT MOD RES 4 4  
FT MOD RES 10 10  
FT MOD RES 10 10  
SQ SEQUENCE 10 AA; 1306 MW; 99DBFCD37861B5A CRC64;

Query Match 22.6%; Score 14; DB 1; Length 10;  
Best Local Similarity 66.7%; Pred. No. 6.6e+03;  
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 RDY 8

Thu Sep 9 07:45:57 2004

us-10-713-208-6\_copy\_184\_193.rsp

Page 6

Do : ||  
2 QDY 4

Search completed: September 9, 2004, 06:21:06  
Job time : 12 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:18:37; Search time 37 Seconds

(without alignments)  
85,275 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193

Perfect score: 62

Sequence: 1 CMPEHRDYS 10

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1443

Minimum DB seq length: 0

Maximum DB seq length: 10

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPREMBL\_25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriophage.\*  
17: sp\_archaeoph.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	32.3	10	2	P83160
2	19	30.6	10	4	Q8N6B1
3	18	29.0	7	12	Q6113
4	18	29.0	8	2	Q9ZIE9
5	18	29.0	9	4	Q99887
6	17	27.4	8	11	Q80XV8
7	17	27.4	9	2	Q8R7H9
8	16	25.8	8	6	Q9GME3
9	16	25.8	8	6	Q8R866
10	16	25.8	8	8	Q35792
11	16	25.8	9	6	Q9GJV2
12	16	25.8	9	6	Q9GJV3
13	16	25.8	9	6	Q9GJV1
14	16	25.8	10	12	Q83398
15	16	25.8	10	13	Q73589
16	16	25.8	10	15	Q86324

17	16	25.8	10	15	Q86325	Q86325 rous sarcom
18	16	25.8	10	15	Q86326	Q86326 rous sarcom
19	15	24.2	10	4	Q86XP4	Q86XP4 homo sapien
20	15	24.2	10	8	Q9ZIE9	Q9ZIE9 rous sarcom
21	15	24.2	10	8	Q9ZIE9	Q9ZIE9 rous sarcom
22	15	24.2	10	11	Q63056	Q63056 rous sarcom
23	14.5	23.4	10	11	Q9QV55	Q9QV55 mus sp. pro
24	14	22.6	7	2	Q8G112	Q8G112 dorilla bu
25	14	22.6	8	4	Q15888	Q15888 homo sapien
26	14	22.6	8	11	Q9QV15	Q9QV15 rous sarcom
27	14	22.6	8	11	Q9QV15	Q9QV15 rous sarcom
28	14	22.6	8	12	Q9WJ33	Q9WJ33 pseudorhab
29	14	22.6	9	2	P83222	P83222 streptomyce
30	14	22.6	9	4	Q15891	Q15891 homo sapien
31	14	22.6	9	10	P82072	P82072 euhadra her
32	14	22.6	9	10	P82429	P82429 nicotiana t
33	14	22.6	9	11	Q99MG3	Q99MG3 mus muscula
34	14	22.6	9	11	Q9QZ48	Q9QZ48 mus muscula
35	14	22.6	9	15	Q12096	Q12096 caprine art
36	14	22.6	9	15	Q12100	Q12100 caprine art
37	14	22.6	9	15	Q12102	Q12102 caprine art
38	14	22.6	9	15	Q12098	Q12098 caprine art
39	14	22.6	9	15	Q12104	Q12104 caprine art
40	14	22.6	10	1	Q9UMM5	Q9UMM5 sulfolobus
41	14	22.6	10	2	Q7WU32	Q7WU32 escherichia
42	14	22.6	10	8	Q9TFU9	Q9TFU9 leuconisc
43	14	22.6	10	10	Q8L7F5	Q8L7F5 hevea brasl
44	14	22.6	10	11	Q80WD9	Q80WD9 rous sarcom
45	13	21.0	7	11	Q8K3H6	Q8K3H6 rous sarcom

## ALIGNMENTS

RESULT 1  
ID P83160 PRELIMINARY; PRT; 10 AA.  
AC P83160;  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)  
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
DE Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RUBISCO  
DE small subunit) (Fragment).  
OS Anabaena sp. (strain L31).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
OX NCBI\_Taxid=29412;  
RN [1]  
RP SEQUENCE.  
RA Apte S.K., Uhlmann E., Schmid R., Altendorf K.;  
RL Submitted (Oct-2001) to Swiss-Pol.  
RC - FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC - CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
CC PHOSPHO-D-GLYCERATE.  
CC - CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC - SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC - SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC - SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.  
DR GO:GO:0016829; F:lyase activity; IEA.  
DR GO:GO:0004497; F:monooxygenase activity; IEA.  
DR GO:GO:0016984; F:ribulose-bisphosphate carboxylase activity; IEA.  
DR GO:GO:0009853; P:photosynthesis; IEA.  
DR GO:GO:0019253; P:photosynthesis; IEA.  
DR GO:GO:0019253; P:photosynthesis; IEA.  
DR GO:GO:0019253; P:photosynthesis; IEA.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration; Lyase;  
KW Oxidoreductase; Monooxygenase; Calvin cycle.  
FT NON TER 10  
SQ SEQUENCE 10 AA; 1322 MW; 8BA4E2DOB13276731 CRC64;

Query Match 32.3%; Score 20; DB 2; Length 10;  
 Best Local Similarity 42.9%; Pred. No. 3.6e+03;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPEHRD 8  
 DB 4 LPEKRRY 10

## RESULT 2

08N6B1 PRELIMINARY; PRT; 10 AA.

AC 08N6B1; 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 D3 Capacitative calcium channel protein Trpl (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Late pregnancy myometrium;  
 RX MEDLINE=22181008; PubMed=12193412;  
 RA Yang M., Gupta A., Shlykov S.G., Corrigan B., Tsujimoto S.,  
 RA Sanborn B.M.,  
 RT "Multiple Trp isoforms implicated in capacitative calcium entry are  
 expressed in human pregnant myometrium and myometrial cells.";  
 RL Biol. Reprod. 67:988-994(2002).  
 DR EMBL; AF483646; AAM57861.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 10 AA; 1336 MW; 0DD0C0B401F40724 CRC64;

Query Match 30.6%; Score 19; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.5e+03;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 HRD 7  
 DB 6 HRD 8

## RESULT 3

066113 PRELIMINARY; PRT; 7 AA.

AC 066113; 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 D3 C-terminal of the viral replicase (Fragment).  
 OS Cherry leaf roll virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;  
 OC Nepovirus.  
 CX NCBI\_TaxID=12615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mainut;  
 RX MEDLINE=96124520; PubMed=8560786;  
 RA Borja M., Sanchez F., Rowhani A., Bruening G., Ponz F.,  
 RT "Long, nearly identical untranslated sequences at the 3' terminal  
 regions of the genomic RNAs of cherry leafroll virus (Mainut  
 strain).";  
 RL Virus Genes 10:245-252(1995).  
 DR EMBL; Z34265; CAA84019.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 7 AA; 800 MW; 7417672EBDCED740 CRC64;

Query Match 29.0%; Score 18; DB 12; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 1e+06;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CMP 3  
 DB 4 CLP 6

## RESULT 4

09ZIE9 PRELIMINARY; PRT; 8 AA.

AC 09ZIE9; 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE Carbamoyl-phosphate synthase subunit B (Fragment).  
 GN CARB.  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 CX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1527;  
 RX MEDLINE=95291461; PubMed=7773412;  
 RA Lawson F.S., Billowes F.M., Dillon J.A.,  
 RT "Organization of carbamoyl-phosphate synthase genes in Neisseria  
 gonorrhoeae includes a large, variable intergenic sequence which is  
 also present in other Neisseria species.";  
 RL Microbiology 141:0-0(10).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1527;  
 RA Brinkman F.S.L., Francis F.M., Dillon J.R.,  
 RT "Complexity of the variable sequence between the carbamoyl-phosphate  
 synthase genes of Neisseria species.";  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF029361; AAC78449.1; -.  
 FT NON TER 8  
 SQ SEQUENCE 8 AA; 988 MW; FA372AB1A4032766 CRC64;

Query Match 29.0%; Score 18; DB 2; Length 8;  
 Best Local Similarity 50.0%; Pred. No. 1e+06;  
 Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPEHRD 7  
 DB 1 MPEKRD 6

## RESULT 5

099887 PRELIMINARY; PRT; 9 AA.

AC 099887; 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE 11 <beta>-HSD2 protein (Fragment).  
 GN 11 <beta>-HSD2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96133030; PubMed=8538347;  
 RX Stewart P.M., Krozowski Z.S., Gupta A., Milford D.V., Howie A.J.,  
 RA Sheppard M.C., Krozowski C.B.,  
 RT "Hypertension in the syndrome of apparent mineralocorticoid excess due  
 to mutation of the 11 beta-hydroxysteroid dehydrogenase type 2 gene.";  
 RL Lancet 347:88-91(1996).

```
DR EMBL: S80133; AAD14324.1; -.
DR GO: GO:0003845; F.11-beta-Hydroxysteroid dehydrogenase activity; NAS.
DR GO: GO:0008212; Pimneralocorticoid metabolism; NAS.
DR NON TER 1 1
SQ SEQUENCE 9 AA; 1020 MW; CERC2B1F5B059C9 CRC64;

Query Match 29.0%; Score 18; DB 4; Length 9;
Best Local Similarity 66.7%; Pred. No. 1e+06;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CMP 3
DB 7 CUP 9

RESULT 6
ID O98XV8 PRELIMINARY; PRT; 8 AA.
AC O80XV8;
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE Tropolon I slow isoform (Fragment).
GN TNSLOW.
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10118;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019373; PubMed=8413291;
R Banerjee-Basu S.; Buonanno A.;
RT "cis-acting sequences of the rat tropoin I slow gene confer tissue-
and development-specific transcription in cultured muscle cells as
well as fiber type specificity in transgenic mice.";
RL Mol. Cell. Biol. 13:7019-7028(1993).
DR EMBL: S66172; AAP13969.1; -.
FT NON TER 8
SQ SEQUENCE 8 AA; 987 MW; FE24540B02CB0766 CRC64;

Query Match 27.4%; Score 17; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 MPE 4
DB 1 MPE 3

RESULT 7
ID O9R7H9 PRELIMINARY; PRT; 9 AA.
AC O9R7H9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Lipoprotein (Fragment).
GN NLPD.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=127;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98083063; PubMed=9422600;
R Martin K., Morin G., Smith A., Nordyke A., Eisenstark A., Golomb M.;
RT "The tryptophanase gene cluster of Haemophilus influenzae type b:
evidence for horizontal gene transfer.";
RL J. Bacteriol. 180:1107-118(1998).
DR EMBL: AF003252; AAB96582.1; -.
FT NON TER 1
SQ SEQUENCE 9 AA; 1152 MW; 35A017673B4412D7 CRC64;
```

```
Query Match 27.4%; Score 17; DB 2; Length 9;
Best Local Similarity 50.0%; Pred. No. 1e+06;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPEH 5
DB 6 IPRH 9

RESULT 8
ID O9GMH3 PRELIMINARY; PRT; 8 AA.
AC O9GMH3;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE Actin (Fragment).
OS Lagenorhynchus obscurus (Dusky dolphin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;
OC Lagenorhynchus.
OX NCBI_TaxId=27611;
RN [1]
RP SEQUENCE FROM N.A.
R Hare M.P., Cipriano F., Palumbi S.R.;
RT "Slow Evolution of Genetic Monophyly in Dolphins: Implications for
Speciation, Systematics and Conservation.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF140833; AAF98686.1; -.
FT NON TER 1
SQ SEQUENCE 8 AA; 962 MW; 5BD1F417740862C0 CRC64;

Query Match 25.8%; Score 16; DB 6; Length 8;
Best Local Similarity 50.0%; Pred. No. 1e+06;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 PEHR 6
DB 5 PRHQ 8

RESULT 9
ID O28866 PRELIMINARY; PRT; 8 AA.
AC O28866;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE Actin protein (Fragment).
GN ACTIN.
OS Megaptera novaeangliae (Humpback whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balenopteridae; Megaptera.
OX NCBI_TaxId=9773;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94285813; PubMed=7912407;
R Palumbi S.R., Baker C.S.;
RT "Contrasting population structure from nuclear intron sequences and
mtDNA of humpback whales.";
RL Mol. Biol. Evol. 11:426-435(1994).
DR EMBL: S73467; AAD14118.1; -.
FT NON TER 1
SQ SEQUENCE 8 AA; 906 MW; 69C866D1F4177408 CRC64;

Query Match 25.8%; Score 16; DB 6; Length 8;
Best Local Similarity 50.0%; Pred. No. 1e+06;
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 PEHR 6
DB 5 PRHQ 8
```

Db 3 PRHQ 6

RESULT 10  
Q35792 PRELIMINARY; PRT; 9 AA.

AC Q35792;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Inside Intron 3 (Fragment).  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Mitochondrion.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D273-10B;  
RA MEDLINE=81069885; PubMed=6254986;  
RA Bonitz S.G., Coruzzi G., Thalerfeld B., Tzagoloff A., Macino G.;  
RT "Assembly of the mitochondrial membrane system: Structure and  
RT nucleotide sequence of the gene coding for subunit 1 of yeast  
RT cytochrome oxidase."  
RL J. Biol. Chem. 255:11927-11941(1980).  
DR EMBL; V00694; CAA24063.1; -;  
DR GO; GO:0005739; C:mitochondrion; IEA.  
KW Mitochondrion.  
FT NON\_TER 1 1  
FT NON\_TER 8 8  
SQ SEQUENCE 8 AA; 1025 MW; 08469334555AA337 CRC64;

Query Match 25.8%; Score 16; DB 8; Length 8;  
Best Local Similarity 50.0%; Pred. No. 1e+06;  
Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RDYD 9  
Db 2 KDYN 5

RESULT 11

Q9GJV2 PRELIMINARY; PRT; 9 AA.

AC Q9GJV2;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Actin (Fragment).  
OS Lagenorhynchus obliquidens (Pacific white-sided dolphin).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
OC Lagenorhynchus.  
OX NCBI\_TaxID=90247;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hare M.P., Cipriano F., Palumbi S.R.;  
RT "Slow Evolution of Genetic Monophyly in Dolphins: Implications for  
RT Speciation, Systematics and Conservation."  
RT Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
RL EMBL; AF140831; AAF98684.1; -;  
DR EMBL; AF140826; AAF98679.1; -;  
DR EMBL; AF140827; AAF98680.1; -;  
DR EMBL; AF140828; AAF98681.1; -;  
DR EMBL; AF140829; AAF98682.1; -;  
DR EMBL; AF140830; AAF98683.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 9 9  
SQ SEQUENCE 9 AA; 1049 MW; 1D0EF417740862C0 CRC64;

Query Match 25.8%; Score 16; DB 6; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1e+06;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PEHR 6  
Db 6 PRHQ 9

RESULT 12  
Q9GJV3 PRELIMINARY; PRT; 9 AA.

AC Q9GJV3;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Actin (Fragment).  
OS Lagenorhynchus obscurus (Dusky dolphin).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
OC Lagenorhynchus.  
OX NCBI\_TaxID=27611;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hare M.P., Cipriano F., Palumbi S.R.;  
RT "Slow Evolution of Genetic Monophyly in Dolphins: Implications for  
RT Speciation, Systematics and Conservation."  
RT Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
RL EMBL; AF140834; AAF98687.1; -;  
DR EMBL; AF140835; AAF98688.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 9 9  
SQ SEQUENCE 9 AA; 1049 MW; 1D0EF417740862C0 CRC64;

Query Match 25.8%; Score 16; DB 6; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1e+06;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PEHR 6  
Db 6 PRHQ 9

RESULT 13

Q9GJV1 PRELIMINARY; PRT; 9 AA.

AC Q9GJV1;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE Actin (Fragment).  
OS Lagenorhynchus acutus (Atlantic white-sided dolphin).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Delphinidae;  
OC Lagenorhynchus.  
OX NCBI\_TaxID=90246;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hare M.P., Cipriano F., Palumbi S.R.;  
RT "Slow Evolution of Genetic Monophyly in Dolphins: Implications for  
RT Speciation, Systematics and Conservation."  
RT Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
RL EMBL; AF140825; AAF98678.1; -;  
DR EMBL; AF140822; AAF98675.1; -;  
DR EMBL; AF140823; AAF98676.1; -;  
DR EMBL; AF140824; AAF98677.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 9 9  
SQ SEQUENCE 9 AA; 1049 MW; 1D0EF417740862C0 CRC64;

Query Match 25.8%; Score 16; DB 6; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1e+06;  
Matches 2; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 6 PRHQ 9

Db :|||  
6 LPEVR 10

## RESULT 14

Search completed: September 9, 2004, 06:21:57  
Job time : 47 secs

Q83978 PRELIMINARY; PRT: 10 AA.  
AC Q83978;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Influenza A/udorn/72 (H3n2), nucleoprotein (Seg 5), 3' cDNA (Fragment).  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83112211; PubMed=6296449;  
RA Lin B.-C., Lai C.-J.;  
RT "The influenza virus nucleoprotein synthesized from cloned dna in a  
RT simian virus 40 vector is detected in the nucleus."  
RL J. Virol. 45:434-438(1983).  
DR EMBL; J02171; AAA43469.1; -;  
DR GO; GO:0019013; C:Viral nucleocapsid; IEA.  
KW Nucleocapsid.  
FT NON\_TER 1  
SQ SEQUENCE 10 AA; 1173 MW; 8787655B1BDD44A CRC64;

## Query Match

25.8%; Score 16; DB 12; Length 10;

Best Local Similarity 50.0%; Pred. No. 2e+04; Mismatches 0; Indels 0; Gaps 0;

Matches 2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYDS 10

:|||  
7 EXDN 10

## RESULT 15

O73589

PRELIMINARY; PRT: 10 AA.

AC O73589;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Orthodenticle-related homeobox protein (Fragment).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=White Leghorn; TISSUE=Embryo;

RX MEDLINE=98141813; PubMed=9473273;

RA Peale F.V. Jr., Mason K., Hunter A.W., Bothwell M.;

RT "Multiplex display polymerase chain reaction amplifies and resolves

RT related sequences sharing a single moderately conserved domain.";

RL Anal. Biochem. 256:158-168(1998).

DR EMBL; U26149; AAC06187.1; -;

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0005677; F:DNA binding; IEA.

KW DNA-binding; Homeobox; Nuclear protein.

FT NON\_TER 1

SQ SEQUENCE 10 AA; 1156 MW; 5C7BCE07672452C3 CRC64;

Query Match 25.8%; Score 16; DB 13; Length 10;

Best Local Similarity 60.0%; Pred. No. 2e+04;

Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPEHR 6

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: September 9, 2004, 06:17:33 ; Search time 52 Seconds

(without alignments)  
54.336 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193

Perfect score: 62  
Sequence: 1 CMPEHRDYDS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 325896

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	46.8	10	7	ADE69558 Human 161
2	29	46.8	10	7	ADE66687 Human 161
3	28	45.2	9	7	ADE66237 Human 161
4	28	45.2	9	7	ADE67499 Human 161
5	28	45.2	9	7	ADE66045 Human 161
6	28	45.2	9	7	ADE66555 Human 161
7	28	45.2	9	7	ADE68727 Human 161
8	28	45.2	9	7	ADE67805 Human 161
9	28	45.2	9	7	ADE67742 Human 161
10	28	45.2	10	2	AAR61945 MBP Pepti
11	28	45.2	10	7	ADE69221 Human 161
12	28	45.2	10	7	ADE70044 Human 161
13	28	45.2	10	7	ADE68695 Human 161
14	27	43.5	8	7	AAR62973 Rg2 Gliom
15	27	43.5	10	2	AAY45153 Monoclonal
16	27	43.5	10	5	ABJ10582 Galatinin-1
17	26	41.9	7	3	AAY87617 Murine G
18	26	41.9	7	3	AAB23207 Hsp47-bln
19	26	41.9	7	3	ADD42081 FP recept
20	26	41.9	8	3	AAY87612 Murine G
21	26	41.9	8	3	AAY87609 Murine G
22	26	41.9	8	3	ADD42076 FP recept
23	26	41.9	8	7	ADD42082 FP recept
24	26	41.9	8	7	ADD42071 Prostagla
25	26	41.9	8	7	ADD42075 FP recept

26	26	41.9	8	7	ADD42073 Prostagla
27	26	41.9	9	4	AAM25058 Human MHC
28	26	41.9	9	4	AAM24935 Human MHC
29	26	41.9	10	2	AAM50212 Sequence
30	26	41.9	10	4	AAM25001 Human MHC
31	26	41.9	10	4	AAM25100 Human MHC
32	25	40.3	6	4	AAG63113 Amino aci
33	25	40.3	6	6	ABU58298 Alpha-1-P
34	25	40.3	7	2	AAY14734 Serine th
35	25	40.3	7	6	AAY14734 Serine th
36	25	40.3	9	2	AAM60376 Tumour ho
37	25	40.3	9	2	AAR93702 Human bre
38	25	40.3	9	3	AAB09440 Hepatic tis
39	25	40.3	9	3	AAB21793 Human bre
40	25	40.3	9	4	AAB06371 Human bre
41	25	40.3	10	5	ABG67536 Human ADP
42	25	40.3	10	6	ADA23652 Alzheimer
43	24	38.7	6	6	ABR43827 A. niger
44	24	38.7	8	2	AAR73361 Human TSH
45	24	38.7	8	2	AAR95749 Alpha-4be

## ALIGNMENTS

RESULT 1	
ID ADE69558	standard; peptide, 10 AA.
AC ADE69558;	
DT 29-JAN-2004	(first entry)
DE Human 161P2F10B protein-related peptide 3563.	
KW 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.	
XX Homo sapiens.	
OS	
PN WO2003040340-A2.	
PD 15-MAY-2003.	
PF 07-NOV-2002; 2002MO-US036002.	
PR 07-NOV-2001; 2001US-00005480.	
PR 31-JAN-2002; 2002US-00062109.	
XX	
PA (AGEN-) AGENSYS INC.	
PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KJM;	
PI Morrison RK, Challita-Eid PM,	
XX WPI; 2003-441560/41.	
XX	
PT A composition for diagnosing, preventing and treating cancer (e.g.	
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides	
PT and polypeptides.	
XX	
PS Claim 13; Page 171; 135pp; English.	
XX	
CC This invention relates to a novel composition which comprises a substance	
CC that modulates the status of a novel human protein (161P2F10B) and its	
CC variants having a sequence of 875 amino acids provided in the	
CC specification. The protein of the invention is over-expressed in certain	
CC cancers. The compounds of the invention may have cytostatic activity and	
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may	
CC be useful for gene therapy or the development of a vaccine. The	
CC composition and methods of the invention are useful in diagnosing,	
CC preventing and treating cancer. The present sequence is the amino acid	
CC sequence of a peptide which is derived from the sequence of the human	
CC 161P2F10B protein and which may be used in the development of the	
CC compounds of the invention.	

XX  
SQ Sequence 10 AA;

Query Match 46.8%; Score 29; DB 7; Length 10;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CMPEHRDYDS 10  
|:|:|:  
1 CLYHREYVS 10

RESULT 2  
ADE66687  
ID ADE66687 standard; peptide; 10 AA.

AC ADE66687;

DT 29-JAN-2004 (first entry)

DE Human 161P2F10B protein-related peptide 694.

KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.

OS Homo sapiens.

PN WO2003040340-A2.

PD 15-MAY-2003.

PF 07-NOV-2002; 2002MO-US036002.

PR 07-NOV-2001; 2001US-00005480.

PR 31-JAN-2002; 2002US-00062109.

PA (AGEN-) AGENSYS INC.

PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KM;

PI Morrison RK, Chailita-Eld PM;

DR WPI; 2003-441560/41.

PT A composition for diagnosing, preventing and treating cancer (e.g.

PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides

and polypeptides.

PS Claim 13; Page 150; 135pp; English.

XX This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.

XX  
SQ Sequence 10 AA;

Query Match 46.8%; Score 29; DB 7; Length 10;  
Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CMPEHRDYDS 10  
|:|:|:  
1 CLYHREYVS 10

RESULT 3

ADE66237  
ID ADE66237 standard; peptide; 9 AA.

XX  
AC ADE66237;  
XX  
DT 29-JAN-2004 (first entry)

DE Human 161P2F10B protein-related peptide 244.

KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.

OS Homo sapiens.

PN WO2003040340-A2.

PD 15-MAY-2003.

PF 07-NOV-2002; 2002MO-US036002.

PR 07-NOV-2001; 2001US-00005480.

PR 31-JAN-2002; 2002US-00062109.

PA (AGEN-) AGENSYS INC.

PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KM;

PI Morrison RK, Chailita-Eld PM;

DR WPI; 2003-441560/41.

PT A composition for diagnosing, preventing and treating cancer (e.g.  
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides  
and polypeptides.

PS Claim 13; Page 147; 135pp; English.

XX This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.

XX  
SQ Sequence 9 AA;

Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CMPEHRDY 8  
|:|:|:  
1 CLYHREY 8

RESULT 4  
ADE67499

ID ADE67499 standard; peptide; 9 AA.

AC ADE67499;

DT 29-JAN-2004 (first entry)

DE Human 161P2F10B protein-related peptide 1504.

KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.

OS Homo sapiens.

XX



PN WO2003040340-A2.  
XX  
PD 15-MAY-2003.  
XX  
PF 07-NOV-2002; 2002WO-US036002.  
XX  
PR 07-NOV-2001; 2001US-00005480.  
XX  
PS 31-JAN-2002; 2002US-00062109.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KJM,  
PI Morrison RK, Challita-Bid PM;  
XX  
DR MPI; 2003-441560/41.  
XX  
PT A composition for diagnosing, preventing and treating cancer (e.g.  
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides  
XX and polypeptides.  
XX  
PS Claim 13; Page 157; 135pp; English.  
XX  
CC This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.  
XX  
SQ Sequence 9 AA;  
XX  
Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CMPEHRDY 8  
|:|:|:  
Db 2 CLVYHREY 9  
XX  
RESULT 5  
ADE66045 ID ADE66045 standard; peptide; 9 AA.  
XX  
AC ADE66045;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human 161P2F10B protein-related peptide 52.  
XX  
KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003040340-A2.  
XX  
PD 15-MAY-2003.  
XX  
PF 07-NOV-2002; 2002WO-US036002.  
XX  
PR 07-NOV-2001; 2001US-00005480.  
XX  
PS 31-JAN-2002; 2002US-00062109.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KJM,  
PI Morrison RK, Challita-Bid PM;

XX  
DR MPI; 2003-441560/41.  
XX  
PT A composition for diagnosing, preventing and treating cancer (e.g.  
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides  
XX and polypeptides.  
XX  
PS Claim 13; Page 145; 135pp; English.  
XX  
CC This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.  
XX  
SQ Sequence 9 AA;  
XX  
Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CMPEHRDY 8  
|:|:|:  
Db 2 CLVYHREY 9  
XX  
RESULT 6  
ADE66555 ID ADE66555 standard; peptide; 9 AA.  
XX  
AC ADE66555;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human 161P2F10B protein-related peptide 562.  
XX  
KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003040340-A2.  
XX  
PD 15-MAY-2003.  
XX  
PF 07-NOV-2002; 2002WO-US036002.  
XX  
PR 07-NOV-2001; 2001US-00005480.  
XX  
PS 31-JAN-2002; 2002US-00062109.  
XX  
PA (AGEN-) AGENSYS INC.  
XX  
PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KJM,  
PI Morrison RK, Challita-Bid PM;  
XX  
DR MPI; 2003-441560/41.  
XX  
PT A composition for diagnosing, preventing and treating cancer (e.g.  
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides  
XX and polypeptides.  
XX  
PS Claim 13; Page 149; 135pp; English.  
XX  
CC This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain

CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.

CC Sequence 9 AA:

Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CMPEHRDY 8  
|: ||: |  
Db 1 CLLYHREY 8

RESULT 7

AD66727 ADE6727 standard; peptide; 9 AA.

AC ADE6727;

DT 29-JAN-2004 (first entry)

DE Human 161P2F10B protein-related peptide 2732.

KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.

OS Homo sapiens.

EN WO2003040340-A2.

PD 15-MAY-2003.

PF 07-NOV-2002; 2002WO-US036002.

PR 07-NOV-2001; 2001US-00005480.

PR 31-JAN-2002; 2002US-00062109.

PA (AGEN-) AGENSYS INC.

PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KM;

PI Morrison RK, Challita-Eid PW;

DR WPI; 2003-441560/41.

XX A composition for diagnosing, preventing and treating cancer (e.g.

PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides

PT and polypeptides.

PS Claim 13; Page 166; 135pp; English.

XX This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.

XX Sequence 9 AA:

Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CMPEHRDY 8  
|: ||: |  
Db 2 CLLYHREY 9

RESULT 8

AD67805 ADE67805 standard; peptide; 9 AA.

AC ADE67805;

DT 29-JAN-2004 (first entry)

DE Human 161P2F10B protein-related peptide 1810.

KM 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.

OS Homo sapiens.

EN WO2003040340-A2.

PD 15-MAY-2003.

PF 07-NOV-2002; 2002WO-US036002.

PR 07-NOV-2001; 2001US-00005480.

PR 31-JAN-2002; 2002US-00062109.

PA (AGEN-) AGENSYS INC.

PI Jakobovits A, Raitano AB, Faris M, Hubert RS, Ge W, Morrison KM;

PI Morrison RK, Challita-Eid PW;

DR WPI; 2003-441560/41.

XX A composition for diagnosing, preventing and treating cancer (e.g.

PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides

PT and polypeptides.

PS Claim 13; Page 159; 135pp; English.

XX This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC the sequence of the 161P2F10B protein, and the gene which encodes it, may  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.

XX Sequence 9 AA:

Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 CMPEHRDY 8  
|: ||: |  
Db 1 CLLYHREY 8

RESULT 9

AD67742 ADE67742 standard; peptide; 9 AA.

AC ADE67742;

XX

DT 29-JAN-2004 (first entry)  
XX  
DB Human 161P2F10B protein-related peptide 1747.  
XX  
KW 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003040340-A2.  
PD 15-MAY-2003.  
XX  
PF 07-NOV-2002; 2002WO-US036002.  
XX  
PR 07-NOV-2001; 2001US-00005480.  
PR 31-JAN-2002; 2002US-00062109.  
XX  
PA (AGEN-) AGENSYS INC.  
PI Jakovcyls A, Raitano AB, Paris M, Hubert RS, Ge W, Morrison KM,  
PI Morrison RK, Challita-Bid PM;  
XX  
DR WPI; 2003-441560/41.  
XX  
PT A composition for diagnosing, preventing and treating cancer (e.g.  
PT prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides  
PT and polypeptides.  
XX  
PS Claim 13; Page 159; 135pp; English.  
XX  
CC This invention relates to a novel composition which comprises a substance  
CC that modulates the status of a novel human protein (161P2F10B) and its  
CC variants having a sequence of 875 amino acids provided in the  
CC specification. The protein of the invention is over-expressed in certain  
CC cancers. The compounds of the invention may have cytostatic activity and  
CC be useful for gene therapy or the development of a vaccine. The  
CC composition and methods of the invention are useful in diagnosing,  
CC preventing and treating cancer. The present sequence is the amino acid  
CC sequence of a peptide which is derived from the sequence of the human  
CC 161P2F10B protein and which may be used in the development of the  
CC compounds of the invention.  
XX  
SQ Sequence 9 AA;  
XX  
Query Match 45.2%; Score 28; DB 7; Length 9;  
Best Local Similarity 50.0%; Pred. No. 1.4e+06;  
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CMPEHRDY 8  
DB 2 CULYHREV 9  
XX  
RESULT 10  
ID AAR61945 standard; peptide; 10 AA.  
XX  
AC AAR61945;  
XX  
DT 25-MAR-2003 (revised)  
DT 31-MAY-1995 (first entry)  
XX  
DE MBP peptide 28, potential binder of HLA-A2.1.  
XX  
XX antigen; epitope; immunogenic target protein; PSA; HBVC; HBVs; EBV; HIV;  
KW plasma specific antigen; hepatitis B virus; Epstein Barr;  
KW human immunodeficiency virus; human papilloma virus; p53; c-ERB2; MAGE-1;  
KW melanoma antigen-1; core antigen; surface antigen;  
KW pharmacokinetic composition; in vivo; ex vivo; therapeutic; diagnostic;  
KW MHC class I molecule; major histocompatibility complex; HLA-A2.1; 9mer;  
KW 10mer; anchor; human leukocyte antigen; PIP; 6mer; algorithm prediction;  
KW MBP.  
XX

XX  
OS Homo sapiens.  
XX  
PN WO9420127-A1.  
XX  
PD 15-SEP-1994.  
XX  
PF 04-MAR-1994; 94WO-US002353.  
XX  
PR 05-MAR-1993; 93US-00027146.  
PR 04-JUN-1993; 93US-00073205.  
PR 29-NOV-1993; 93US-00159184.  
XX  
PA (CYTE-) CYTEL CORP.  
PI Grey HM, Settle A, Sidney J, Kast W;  
PI WPI; 1994-302678/37.  
XX  
DR Immunogenic peptide(s) having an HLA-A2.1 binding motif - used for  
PT treatment or prophylaxis of cancer, virus infection or autoimmune  
PT diseases.  
XX  
PS Disclosure; Page 128; 138pp; English.  
XX  
CC AAR61714-837 are potential peptide binders of HLA-A2.1 motif. These  
CC peptides are thus potentially immunogenic. They were predicted by using  
CC an algorithm, which assigns a score for each amino acid, at each position  
CC along a peptide. A peptide is scored in the 'Grouped Ratio' algorithm as  
CC a product of the scores of each of its residues. This peptide has an  
CC algorithm score (EO2) of -27.85. This value can then be used to predict a  
CC population of peptides with the highest occurrence of good binders. The  
CC peptides of the invention can induce cytotoxic T lymphocytes which can  
CC react with target cells. They can be used for the treatment or  
CC prophylaxis of cancer, eg. prostate cancer or lymphoma, etc. (updated on  
CC 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 10 AA;  
XX  
Query Match 45.2%; Score 28; DB 2; Length 10;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 MPEHRD 7  
DB 2 LPRHRD 7  
XX  
RESULT 11  
ID ADE69221 standard; peptide; 10 AA.  
XX  
AC ADE69221;  
XX  
DT 29-JAN-2004 (first entry)  
DE Human 161P2F10B protein-related peptide 3226.  
XX  
KW 161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.  
XX  
OS Homo sapiens.  
XX  
PN WO2003040340-A2.  
PD 15-MAY-2003.  
XX  
XX 07-NOV-2002; 2002WO-US036002.  
PF 07-NOV-2001; 2001US-00005480.  
PR 31-JAN-2002; 2002US-00062109.  
XX  
PA (AGEN-) AGENSYS INC.  
XX

```

CC variants having a sequence of 875 amino acids provided in the
CC specification. The protein of the invention is over-expressed in certain
CC cancers. The compounds of the invention may have cyclostatic activity and
CC the sequence of the 1612PFI0B protein, and the gene which encodes it, may
CC be useful for gene therapy or the development of a vaccine. The
CC composition and methods of the invention are useful in diagnosing,
CC preventing and treating cancer. The present sequence is the amino acid
CC sequence of a peptide which is derived from the sequence of the human
CC 1612PFI0B protein and which may be used in the development of the
CC compounds of the invention.
XX
SQ Sequence 10 AA;

Query Match          45.2%; Score 28; DB 7; Length 10;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0

QY 1 CMPEHRDY 8
   | : | : |
Db 3 CLLYHREY 10

RESULT 13
AD569695
ID AD569695 standard; peptide; 10 AA.

```

```

QY      1 CMPEHRDY 8
      | : | | : |
Db      3 CLLYHREY 10

RESULT 13
ADE69695
1D      ADE69695 standard; peptide; 10 AA.
XX
XX      ADE69695;
XX
XX      29-JAN-2004 (first entry)
XX
DE      Human 161P2F10B protein-related peptide 3700.
XX
XX      161P2F10B; cancer; cytostatic; gene therapy; vaccine; human.
XX

```

XX	WO2003040340-A2.
PN	
XX	
PD	15-MAY-2003.
XX	
PF	07-NOV-2002; 2002WO-US036002.
XX	
XX	07-NOV-2001; 2001US-00005480.
PR	31-JAN-2002; 2002US-00062109.
XX	
PA	(AGEN-) AGENSYS INC.
XX	
PI	Jakobovits A, Raitano AB, Paris M, Hubert RS, Ge W, Morrison KJM;
XX	Morrison RK, Chailita-Bid PM;
XX	
DR	WPI; 2003-441560/41.
XX	
PT	A composition for diagnosing, preventing and treating cancer (e.g.
XX	prostatic, renal or uterine cancer) comprises 161P2F10B polynucleotides
PT	and polypeptides.
XX	
PS	Claim 13; Page 172; 135pp; English.
CC	
CC	This invention relates to a novel composition which comprises a substance
CC	that modulates the status of a novel human protein (161P2F10B) and its
CC	variants having a sequence of 875 amino acids provided in the
CC	specification. The protein of the invention is over-expressed in certain
CC	cancers. The compounds of the invention may have cytostatic activity and
CC	the sequence of the 161P2F10B protein, and the gene which encodes it, may
CC	be useful for gene therapy or the development of a vaccine. The
CC	composition and methods of the invention are useful in diagnosing,
CC	preventing and treating cancer. The present sequence is the amino acid
CC	sequence of a peptide which is derived from the sequence of the human
CC	161P2F10B protein and which may be used in the development of the
CC	compounds of the invention.
XX	
XX	Sequence 10 AA;
5Q	



*This Page Blank (uspro)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:21:02 (Search time 46 Seconds  
(without alignments)  
69.715 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_184\_193

Perfect score: 62  
Sequence: 1 CMPEHRDYDS 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 160061

Minimum DB seq length: 0  
Maximum DB seq length: 10

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	46.8	10	US-10-094-401-170	Sequence 170, App
2	29	46.8	10	US-10-462-262-138	Sequence 138, App
3	28	45.2	9	US-10-062-109A-102	Sequence 102, App
4	28	45.2	9	US-10-005-480A-102	Sequence 102, App
5	27	43.5	8	US-10-357-929A-5	Sequence 5, Appl
6	27	43.5	10	US-10-466-205-4	Sequence 44, Appl
7	26	41.9	9	US-09-780-053-428	Sequence 428, App
8	26	41.9	9	US-09-780-053-551	Sequence 551, App
9	26	41.9	10	US-09-780-053-494	Sequence 494, App
10	26	40.3	9	US-09-780-053-593	Sequence 593, App
11	25	40.3	6	US-09-727-963A-52	Sequence 52, Appl
12	25	40.3	9	US-08-424-550B-567	Sequence 567, Appl
13	25	40.3	9	US-09-765-086-93	Sequence 93, Appl
14	25	40.3	9	US-10-264-374-93	Sequence 93, Appl
15	25	40.3	14	US-10-375-992-93	Sequence 93, Appl

15	25	40.3	9	16	US-10-264-374-93	Sequence 93, Appl
16	25	40.3	9	16	US-10-375-992-93	Sequence 93, Appl
17	25	40.3	10	12	US-10-014-340-262	Sequence 262, Appl
18	25	40.3	10	8	US-08-964-716-2	Sequence 2, Appl
19	24	38.7	10	10	US-09-572-404B-1163	Sequence 1163, App
20	24	38.7	10	10	US-09-572-404B-3392	Sequence 3392, Appl
21	24	38.7	10	10	US-09-573-822C-510	Sequence 510, Appl
22	24	38.7	10	10	US-09-573-822C-512	Sequence 512, Appl
23	24	38.7	7	9	US-09-884-767A-49	Sequence 49, Appl
24	23	37.1	8	15	US-10-357-929A-21	Sequence 21, Appl
25	23	37.1	9	9	US-09-834-765-226	Sequence 226, App
26	23	37.1	9	9	US-09-834-765-316	Sequence 316, App
27	23	37.1	9	9	US-09-834-765-657	Sequence 657, App
28	23	37.1	9	9	US-09-834-765-657	Sequence 657, App
29	23	37.1	9	10	US-09-741-744A-126	Sequence 126, Appl
30	23	37.1	10	9	US-09-819-308-24	Sequence 24, Appl
31	23	37.1	10	9	US-09-834-765-95	Sequence 95, Appl
32	23	37.1	10	9	US-09-834-765-189	Sequence 189, App
33	23	37.1	10	9	US-09-834-765-189	Sequence 297, App
34	23	37.1	10	9	US-09-834-765-373	Sequence 373, App
35	23	37.1	10	9	US-09-834-765-492	Sequence 492, App
36	23	37.1	10	9	US-09-834-765-512	Sequence 512, App
37	23	37.1	10	9	US-09-834-765-572	Sequence 572, App
38	23	37.1	10	13	US-10-054-552-2	Sequence 2, Appl
39	23	37.1	10	14	US-10-062-109A-65	Sequence 65, Appl
40	23	37.1	10	14	US-10-062-109A-349	Sequence 349, Appl
41	23	37.1	10	14	US-10-062-109A-404	Sequence 404, Appl
42	23	37.1	10	14	US-10-005-480A-65	Sequence 65, Appl
43	23	37.1	10	14	US-10-005-480A-49	Sequence 349, Appl
44	23	37.1	10	14	US-10-005-480A-404	Sequence 404, Appl
45	23	37.1	10	15	US-10-137-867-350	Sequence 390, Appl

## ALIGNMENTS

```
RESULT 1
US-10-094-401-170
; Sequence 170, Application US/10094401
; Publication No. US20030069395A1
; GENERAL INFORMATION:
; APPLICANT: DYAX CORP.
; APPLICANT: Dyax, Aaron K.
; APPLICANT: Ley, Arthur C.
; APPLICANT: Cohen, Edward H.
; TITLE OF INVENTION: SERUM ALBUMIN BINDING MOETITIES
; FILE REFERENCE: DYX-026.2 PCT; DYX-026.2 US
; CURRENT APPLICATION NUMBER: US/10/094,401
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 60/331,352
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/292,975
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 271
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 170
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: albumin binding peptide
US-10-094-401-170
```

Query Match 46.8% Score 29; DB 14; Length 10;  
Best Local Similarity 66.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1 CM-PEHRD 7  
|||  
DB 1 CMDPEHRD 9

RESULT 2  
US-10-462-262-138

```
; Sequence 138, Application US/10462262
; Publication No. US20040009534A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Aaron K.
; APPLICANT: Dawson, Bruce M.
; TITLE OF INVENTION: PROTEIN ANALYSIS
; FILE REFERENCE: 10280-052001
; CURRENT APPLICATION NUMBER: US/10/462,262
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: US 60/388,642
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 430
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: example of serum albumin-binding agents
US-10-462-262-138
```

```
Query Match          46.8%; Score 29; DB 15; Length 10;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 2; Gaps 1;
```

```
QY 1 CM--PEHRD 7
Db 1 CNDWPHRD 9
```

```
RESULT 3
US-10-062-109A-102
; Sequence 102, Application US/10062109A
; Publication No. US20030165505A1
; GENERAL INFORMATION:
; APPLICANT: Agensys
; APPLICANT: Chailita-Eld, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: Nucleic Acid and Corresponding Protein
; TITLE OF INVENTION: Entitled 161P2F10B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20062.01
; CURRENT APPLICATION NUMBER: US/10/062,109A
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 10/005,480
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-109A-102
```

```
Query Match          45.2%; Score 28; DB 14; Length 9;
Best Local Similarity 50.0%; Pred. No. 1.2e+06;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 CMPEHRDY 8
Db 1 CLYHREY 8
```

```
RESULT 4
US-10-005-480A-102
; Sequence 102, Application US/10005480A
; Publication No. US20030191073A1
; GENERAL INFORMATION:
; APPLICANT: Agensys
```

```
; APPLICANT: Chailita-Eld, Pia M.
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Paris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Morrison, Karen Jane Meyrick
; APPLICANT: Jakobovits, Aya
; TITLE OF INVENTION: Nucleic Acid and Corresponding Protein
; TITLE OF INVENTION: Entitled 161P2F10B Useful in Treatment and Detection of
; FILE REFERENCE: 51158-20062.00
; CURRENT APPLICATION NUMBER: US/10/005,480A
; CURRENT FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 765
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-005-480A-102
```

```
Query Match          45.2%; Score 28; DB 14; Length 9;
Best Local Similarity 50.0%; Pred. No. 1.2e+06;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 CMPEHRDY 8
Db 1 CLYHREY 8
```

```
RESULT 5
US-10-357-929A-5
; Sequence 5, Application US/10357929A
; Publication No. US20030216322A1
; GENERAL INFORMATION:
; APPLICANT: Tatiana I. Samoylova
; APPLICANT: Valery A. Petroenko
; APPLICANT: Nancy R. Cox
; APPLICANT: Nancy E. Morrison
; APPLICANT: Henry J. Baker
; APPLICANT: Ludmila P. Globa
; TITLE OF INVENTION: Peptides for Recognition and Targeting
; TITLE OF INVENTION: of Gli1 Cell Tumors
; FILE REFERENCE: 35721/259047
; CURRENT APPLICATION NUMBER: US/10/357,929A
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 10/357,929
; PRIOR FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 60/354,188
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-10-357-929A-5
```

```
Query Match          43.5%; Score 27; DB 15; Length 8;
Best Local Similarity 80.0%; Pred. No. 1.2e+06;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 MPEHR 6
Db 3 LPEHR 7
```

```
RESULT 6
US-10-466-205-44
; Sequence 44, Application US/10466205
; Publication No. US20040077531A1
; GENERAL INFORMATION:
```



```

; APPLICANT: MATSUMOTO, Hirokazu
; APPLICANT: NOGUCHI, Jiro
; APPLICANT: OHYAKI, Tetsuya
; TITLE OF INVENTION: Use of Galatin-like Peptide
; FILE REFERENCE: 2861USOP
; CURRENT APPLICATION NUMBER: US/10/466,205
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: PCT/JP02/00313
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: JP 2001-12094
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 44
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Porcine
US-10-466-205-44
```

```

Query Match          43.5%; Score 27; DB 16; Length 10;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 CMEPR 6
Db      1 CAPAR 6
```

```

RESULT 7
US-09-780-053-428
; Sequence 428, Application US/09780053
; Patent No. US20020102640A1
; GENERAL INFORMATION:
; APPLICANT: Rene S. Hubert
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Pia M. Challita-Eid
; APPLICANT: Mary Farris
; APPLICANT: Elana Levin
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 83P5G4: A TISSUE SPECIFIC PROTEIN
; FILE REFERENCE: 129.5USU1
; CURRENT APPLICATION NUMBER: US/09/780,053
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/181,261
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 716
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 428
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-780-053-428
```

```

Query Match          41.9%; Score 26; DB 9; Length 9;
Best Local Similarity 80.0%; Pred. No. 1.2e+06;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CMEPR 5
Db      1 CGPRH 5
```

```

RESULT 8
US-09-780-053-551
; Sequence 551, Application US/09780053
; Patent No. US20020102640A1
; GENERAL INFORMATION:
; APPLICANT: Rene S. Hubert
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Pia M. Challita-Eid
; APPLICANT: Mary Farris
; APPLICANT: Elana Levin
```

```

; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 83P5G4: A TISSUE SPECIFIC PROTEIN
; FILE REFERENCE: 129.5USU1
; CURRENT APPLICATION NUMBER: US/09/780,053
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/181,261
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 716
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 551
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-780-053-551
```

```

Query Match          41.9%; Score 26; DB 9; Length 9;
Best Local Similarity 80.0%; Pred. No. 1.2e+06;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CMEPR 5
Db      1 CGPRH 5
```

```

RESULT 9
US-09-780-053-494
; Sequence 494, Application US/09780053
; Patent No. US20020102640A1
; GENERAL INFORMATION:
; APPLICANT: Rene S. Hubert
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Pia M. Challita-Eid
; APPLICANT: Mary Farris
; APPLICANT: Elana Levin
; APPLICANT: Steve Chappell Mitchell
; APPLICANT: Aya Jakobovits
; TITLE OF INVENTION: 83P5G4: A TISSUE SPECIFIC PROTEIN
; FILE REFERENCE: 129.5USU1
; CURRENT APPLICATION NUMBER: US/09/780,053
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/181,261
; PRIOR FILING DATE: 2000-02-09
; NUMBER OF SEQ ID NOS: 716
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 494
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-780-053-494
```

```

Query Match          41.9%; Score 26; DB 9; Length 10;
Best Local Similarity 80.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 CMEPR 5
Db      2 CGPRH 6
```

```

RESULT 10
US-09-780-053-593
; Sequence 593, Application US/09780053
; Patent No. US20020102640A1
; GENERAL INFORMATION:
; APPLICANT: Rene S. Hubert
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Pia M. Challita-Eid
; APPLICANT: Mary Farris
; APPLICANT: Elana Levin
; APPLICANT: Steve Chappell Mitchell
```

APPLICANT: Aya Jakobovits  
TITLE OF INVENTION: 8326G4: A TISSUE SPECIFIC PROTEIN  
TITLE OF INVENTION: HIGHLY EXPRESSED IN PROSTATE CANCER  
FILE REFERENCE: 129, 50501  
CURRENT APPLICATION NUMBER: US/09/780,053  
CURRENT FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/181,261  
PRIOR FILING DATE: 2000-02-09  
NUMBER OF SEQ ID NOS: 716  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 593  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-780-053-593

Query Match 41.3%; Score 26; DB 9; Length 10;  
Best Local Similarity 80.0%; Pred. No. 6.8e+02;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CMPEH 5  
Db 2 CGPEH 6

RESULT 11  
US-09-727-963A-52  
Sequence 52, Application US/09727963A  
Patent No. US20020155106A1  
GENERAL INFORMATION:  
APPLICANT: V.I. Technologies, Inc.  
TITLE OF INVENTION: METHOD OF IDENTIFYING A LIGAND FOR A TARGET MOLECULE  
FILE REFERENCE: 18242-505  
CURRENT APPLICATION NUMBER: US/09/727,963A  
CURRENT FILING DATE: 2000-12-01  
PRIOR APPLICATION NUMBER: U.S.S.N. 60/327,557  
PRIOR FILING DATE: 1999-12-02  
NUMBER OF SEQ ID NOS: 99  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 52  
LENGTH: 6  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Peptide Ligand  
US-09-727-963A-52

Query Match 40.3%; Score 25; DB 9; Length 6;  
Best Local Similarity 80.0%; Pred. No. 1.2e+06;  
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RDYDS 10  
Db 1 RDYDT 5

RESULT 12  
US-08-424-550B-567  
Sequence 567, Application US/08424550B  
Publication No. US20020119447A1  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMT J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAMSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUEHROFF  
APPLICANT: JAMES C. BRKER  
APPLICANT: SHERI L. BUIJK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS

TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/APed  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550B  
FILING DATE:  
CLASSIFICATION: 435435  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 567:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-424-550B-567

Query Match 40.3%; Score 25; DB 8; Length 9;  
Best Local Similarity 66.7%; Pred. No. 1.2e+06;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 EHRDYP 9  
Db 2 QHRDRD 7

RESULT 13  
US-09-765-086-93  
Sequence 93, Application US/09765086  
Patent No. US20010046498A1  
GENERAL INFORMATION:  
APPLICANT: Ruoslahti, Erkki  
APPLICANT: Pasqualini, Renata  
APPLICANT: Madh, Arap  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Ellnerby, H. Michael  
TITLE OF INVENTION: Chimeric Prostate-Homing Peptides With  
Pro-Apoptotic Activity  
FILE REFERENCE: P-LD 3844  
CURRENT APPLICATION NUMBER: US/09/765,086  
CURRENT FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: US 09/489,582  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 235  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 93  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic peptide  
US-09-765-086-93

Query Match 40.3%; Score 25; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PEHR 6  
 ||||  
 Db 2 PEHR 5

RESULT 14  
 US-10-264-374-93  
 ; Sequence 93, Application US/10264374  
 ; Publication No. US20030113320A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruoslahti, Erkki  
 ; APPLICANT: Pasqualini, Renata  
 ; TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing  
 ; TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using  
 ; FILE REFERENCE: P-LJ 3203  
 ; CURRENT APPLICATION NUMBER: US/10/264,374  
 ; CURRENT FILING DATE: 2002-10-03  
 ; PRIOR APPLICATION NUMBER: US/09/139,802  
 ; PRIOR FILING DATE: 1998-08-25  
 ; PRIOR APPLICATION NUMBER: 08/926,914  
 ; PRIOR FILING DATE: 1997-09-10  
 ; PRIOR APPLICATION NUMBER: 08/710,067  
 ; PRIOR FILING DATE: 1996-09-10  
 ; NUMBER OF SEQ ID NOS: 226  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 93  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: Peptide  
 US-10-264-374-93

Query Match 40.3%; Score 25; DB 14; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PEHR 6  
 ||||  
 Db 2 PEHR 5

RESULT 15  
 US-10-375-992-93  
 ; Sequence 93, Application US/10375992  
 ; Publication No. US20030152578A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruoslahti, Erkki  
 ; APPLICANT: Pasqualini, Renata  
 ; TITLE OF INVENTION: Tumor Homing Molecules, Conjugates  
 ; TITLE OF INVENTION: Derived Therefrom, and Methods of Using Same  
 ; NUMBER OF SEQUENCES: 199  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell & Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 92122  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/375,992  
 ; FILING DATE: 27-Feb-2003  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/926,914  
 ; FILING DATE: 10-SEP-1997

ATTORNEY/AGENT INFORMATION:  
 ; NAME: Campbell, Cathryn A.  
 ; REGISTRATION NUMBER: 31,815  
 ; REFERENCE/DOCKET NUMBER: P-LJ 2725  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 535-9001  
 ; TELEFAX: (619) 535-8949  
 ; INFORMATION FOR SEQ ID NO: 93:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: both  
 ; MOLECULE TYPE: peptide  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 93:  
 US-10-375-992-93

Query Match 40.3%; Score 25; DB 14; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PEHR 6  
 ||||  
 Db 2 PEHR 5

Search completed: September 9, 2004, 06:23:24  
 Job time: 47 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:22:39 ; Search time 15 seconds

(without alignments)  
878.549 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_205\_341

Perfect score: 725

Sequence: 1 QSVGVQDQTHSGPLPHAIR.....GTIPGSGITESKDMHFSLG 137

Scoring table: BLOSUM62

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 63571

Minimum DB seq length: 0  
Maximum DB seq length: 137

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	8.3	87	2	AC324
2	59.5	8.2	129	2	H84749
3	59.5	8.2	131	2	T52597
4	56	7.7	94	2	T38652
5	56	7.7	100	2	E69225
6	54.5	7.5	101	2	D82808
7	54.5	7.5	117	2	S47178
8	54.5	7.5	125	2	H69292
9	54	7.4	136	2	I53300
10	53.5	7.4	69	2	A30792
11	53.5	7.4	124	2	T44571
12	53	7.3	119	2	G70367
13	52.5	7.2	129	2	G64508
14	52	7.2	117	2	T24264
15	52	7.2	135	2	S38386
16	51.5	7.1	110	2	PV0006
17	51.5	7.1	116	1	S20630
18	51.5	7.1	135	2	B97301
19	51	7.0	102	2	A72739
20	51	7.0	106	2	S35573
21	51	7.0	137	2	T47418
22	50.5	7.0	75	2	F64423
23	50.5	7.0	102	2	S26346
24	50.5	7.0	109	2	C84525
25	50.5	7.0	116	1	RIHUS1
26	50.5	7.0	116	1	A28968
27	50.5	7.0	123	2	S35479
28	50.5	7.0	137	2	T49433
29	50	6.9	34	2	S44828

30	50	6.9	82	2	AC3437	SSU ribosomal prot
31	50	6.9	111	2	B72482	hypothetical prote
32	50	6.9	121	2	B87679	chemotaxis protein
33	50	6.9	129	2	E82143	conserved hypochet
34	50	6.9	131	2	S65157	probable membrane
35	49.5	6.8	88	2	G96011	hypothetical prote
36	49.5	6.8	103	2	S19975	ig kappa chain V r
37	49.5	6.8	119	2	A96508	hypothetical prote
38	49.5	6.8	127	1	A46528	phosphatidylinosit
39	49.5	6.8	130	2	T45533	gene Thb protein -
40	49.5	6.8	133	2	B32223	interleukin-7 prec
41	49	6.8	74	2	AG2533	hypothetical prote
42	49	6.8	85	2	F66389	probable DNA-bindi
43	49	6.8	118	2	F72411	conserved hypochet
44	49	6.8	124	2	S13159	profilin - slime m
45	49	6.8	125	2	B35273	profilin P - slime

## ALIGNMENTS

RESULT 1  
AC324  
hypothetical protein asl14146 [imported] - Nostoc sp. (strain PCC 7120)  
C/Species: Nostoc sp. PCC 7120  
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C/Accession: AC324  
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriouchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Yasuda, M.; Tabata, D.N. Res. 8, 205-213, 2001  
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A/Reference number: AB1807; MUID:21595285; PMID:11759840  
A/Accession: AC324  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-87 <NUP>  
A/Cross-References: GB:BA000019; PIDN:BA075845.1; PID:G17133281; GSPDB:GN00139  
A/Experimental source: strain PCC 7120  
C/Genetics:  
A/Gene: asl14146

Query Match 8.3%; Score 60.5; DB 2; Length 87;  
Best Local Similarity 29.9%; Pred. NO. 26;  
Matches 20; Conservative 8; Mismatches 24; Indels 15; Gaps 2;  
QY 13 LPLHHRIRMFWDSCPYLAGKPKMFFIONVYVSDQLSDSLLEVDPAMKNVEFKAKR 72  
Db 7 IPUNGIRAPLRGRDPY-----KVQALMESIAATGQCEPDIIVLEVDG-----QYV 51

QY 73 GLCTVHR 79  
Db 52 GFGGCHR 58

RESULT 2  
H84749  
probable squamosa-promoter binding protein [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C/Accession: H84749  
R/Jin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Neus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, A. Nature 402, 761-766, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: AB4420; MUID:20083487; PMID:10617197  
A/Accession: H84749  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-129 <STO>  
A/Cross-References: GB:AE002093; NID:G1707009; PIDN:AA069133.1; GSPDB:GN00139  
C/Genetics:



A:Gene: XF0417

Query Match 7.5%; Score 54.5; DB 2; Length 101;  
Best Local Similarity 20.4%; Pred. No. 1.3e+02;  
Matches 21; Conservative 22; Mismatches 31; Indels 29; Gaps 4;

QY 51 DSSLEVDGPMKNV-----EFKQKRGGLCTVREADFWSLCTADNSLIEGSHSPSL 104  
DB 10 DGNALVLSGPLNRDAIGLMSKICQASQSN-----GVQCLNAGIKQLDSTGVA 58  
QY 105 YVQCLSGKROBGRGT-----PGSGITE-----SKDMHFS 135  
DB 59 LLAEWERIRVOGGSVPTVAPPSGLKEILAAVMSDLPQFS 101

RESULT 7  
S47178  
hypothetical protein pHLA-28 - common tobacco (fragment)  
C/Species: Nicotiana tabacum (common tobacco)  
C/Date: 06-Jan-1995 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C/Accession: S47178  
R/PortJuk, N.; Stetilo, L.  
submitted to the EMBL Data Library, June 1994  
A/Reference number: S47178  
A/Accession: S47178  
A:Molecule type: mRNA  
A/Residues: 1-117 <BOB>  
A/Cross-references: EMBL:X79793, NID:9506468, PID:9506469  
A:Experimental source: cultivar SRI

Query Match 7.5%; Score 54.5; DB 2; Length 117;  
Best Local Similarity 26.6%; Pred. No. 1.6e+02;  
Matches 21; Conservative 18; Mismatches 31; Indels 9; Gaps 4;

QY 13 LPLH-HIRFMWDCPYLAGKKMFQIONVYVSDGLESDSLEVDGPMKNVEFKAQK 71  
DB 6 LPEHEHAKQKF-----EYLQGEAEGLERLHQCAEKELQ--TYLNAGPSKDFENDFTKL 59  
QY 72 RGLCTVREADFWSLCTA 90  
DB 60 AGLTSVTR--NIFENLVRA 76

RESULT 8  
H69292  
desulfoferredoxin homolog - Archaeoglobus fulgidus  
C/Species: Archaeoglobus fulgidus  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 03-Nov-2003  
C/Accession: H69292  
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A/Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attlich, P.; Kaine, B.F.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A/Reference number: A69250; MUID:98049343; PMID:9389475  
A/Accession: H69292  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A/Residues: 1-125 <KLB>  
A/Cross-references: GB:AE001081, GB:AE000782, NID:92689404, PID:AA890892.1, PID:9265029  
C/Superfamily: superoxide reductase/superoxide dismutase; desulfoferredoxin homology

Query Match 7.5%; Score 54.5; DB 2; Length 125;  
Best Local Similarity 23.7%; Pred. No. 1.7e+02;  
Matches 22; Conservative 9; Mismatches 29; Indels 33; Gaps 5;

QY 5 GVDQHGSLPLHHRFM-----GDSCPYLAGKKMFQIONVYVSDGLESDSLEVDG 59  
DB 35 GEIHPNPTTEHIMILVFOPEGSKPPYVVGAAEF-----AAHG-----ASVDG 80  
QY 60 PAMKIV-----EFKAQKRG-----LCIVH 78

DB 81 PNTSGVYTDVAVFAFAKAKSGKLTAFKYNH 113

RESULT 9  
I53300  
interleukin-1-beta-converting enzyme - rat (fragment)  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Nov-1999  
C/Accession: I53300  
R/Plaw, J.A.; Kugu, K.; Trbovich, A.M.; Desanti, A.; Tilly, K.I.; Hirschfield, A.N.; Ti  
Enocrinology 136, 5042-5053, 1995  
A/Title: Interleukin-1-beta-converting enzyme-related proteases (IRPs) and mammalian ce  
nula cells of the ovarian follicle.  
A/Reference number: I53300; MUID:96042508; PMID:7588240  
A/Accession: I53300  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A/Residues: 1-136 <RES>  
A/Cross-references: EMBL:U34621, NID:91002918, PID:AA052259.1, PID:91002919

Query Match 7.4%; Score 54; DB 2; Length 136;  
Best Local Similarity 24.5%; Pred. No. 2.2e+02;  
Matches 23; Conservative 17; Mismatches 44; Indels 10; Gaps 5;

QY 27 CPYLAGKKMFQIONVYVSDGLESDSLEVDGPMKNVEFKA---QKRGGLCTVREAD 82  
DB 42 CPSLDKRKVILIQ--ACRGEKQGVLLK-DSVGNSEEGFLTDALPEDDGIKKAHTEKD 97  
QY 83 FFWSLCTADMSLLEQSHS-SPSLVQCLSGKLRQ 115  
DB 98 FI-ARCSSTPNDVSRHPVQSLFESLIRKMK 130

RESULT 10  
A30792  
calcium-binding protein - fluke (Schistosoma mansoni)  
C/Species: Schistosoma mansoni  
C/Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 13-Aug-1999  
C/Accession: A30792  
R/Ram, D.; Groszman, Z.; Markovics, A.; Avivi, A.; Ziv, E.; Lantner, F.; Schachter, I.  
submitted to GenBank, November 1988  
A/Reference number: A34527  
A/Accession: A30792  
A:Molecule type: mRNA  
A/Residues: 1-69 <RAM>  
A/Cross-references: EMBL:J04641, NID:9160938, PID:AAA29860.1, PID:9160939  
C/Superfamily: calmodulin; calmodulin repeat homology  
C/Keywords: calcium binding; EF hand  
F/2-34/Domain: calmodulin repeat homology <EF1>  
F/38-69/Domain: calmodulin repeat homology <EF2>

Query Match 7.4%; Score 53.5; DB 2; Length 69;  
Best Local Similarity 26.2%; Pred. No. 1.1e+02;  
Matches 17; Conservative 10; Mismatches 25; Indels 13; Gaps 1;

QY 6 VDQFSGSLPLHHRFMFGD-----SCPYLAGKKMFQIONVYVSDGLESDS 52  
DB 2 VNRTFAQLLGLHGRDSKSGISQGLMEFLHTVNCPPKKEQVKEFTKQHKDGDQLNTD 61  
QY 53 SLLEY 57  
DB 62 ELIDV 66

RESULT 11  
T44571  
superoxide reductase (EC 1.-.-.-) [imported] - Pyrococcus furiosus  
C/Species: Pyrococcus furiosus  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 03-Nov-2003  
C/Accession: T44571  
R/Jenney Jr., F.E.; Verhagen, M.F.; Cui, X.; Adams, M.W.  
Science 286, 306-309, 1999





```

QY      5  GVDQTHSL--PLHHI--RMMFNGDSCPYLAGKPKMFFIQ-----NYVSDGOL 49
Db      7  GADHADTVSONPRHKITRGQNTVFRCDPISHHNRLYWRQTLGGQPEFLTYFONEAOL 66
QY     50  EDSJLEVDGPAMKVEFKACK-RGLCTVHREADPFWSLCTADMSLLEQSHSPSLYLQOC 108
Db     67  EKSRLI-----SDRFSABRPKC-----SLSTLEIQRTQGDSD--AMYL-C 103
QY     109  LSQKLRQERGTI-----PSSGITESKDM 131
Db     104  ASTL---RGTSTDTQYFGPGTRITVLEBDL 129

```

Search completed: September 9, 2004, 06:25:45  
 Job time : 16 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:21:13 | Search time 10 Seconds

(without alignments)  
713.361 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_205\_341

Perfect score: 725

Sequence: 1 QGVYGVDPQHSGLPLHHR.....GTFPGSITESKDMHFFSLG 137

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 25926

Minimum DB seq length: 0

Maximum DB seq length: 137

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	ID	Description
1	56	7.7	94	YIP3_SCHPO	G9t66 schizosacch
2	56	7.7	113	HYB1_KLEBN	G9f04 klebsiella
3	56	7.7	124	Y024_PSEBK	G98n8 pseudomonas
4	54.5	7.5	125	S08_TACFU	G9993 archaeoglob
5	53.5	7.4	69	CABP_SCHMA	P1356 schistosoma
6	53.5	7.4	124	S08_PYRFU	P2385 pyrococcus
7	52.5	7.2	129	Y073_METYA	G9967 methanococ
8	52	7.2	113	HYB1_MORMO	G9zh1 moranella
9	51.5	7.1	116	SMS_CHICK	P3304 gallus gall
10	51	7.0	106	RPO1_ABPPE	G9yex1 aeropyrum p
11	51	7.0	106	FRP2_CAUCR	G9839 caulobacter
12	50.5	7.0	75	Y990_METUA	P0166 homo sapien
13	50.5	7.0	116	SMS_HUMAN	G9668 ovis aries
14	50.5	7.0	116	SMS_SHEEP	G9668 ovis aries
15	50.5	7.0	121	GLUC_CARAU	P9695 carassius a
16	50	6.9	82	R818_BRUME	G9yfn9 bruceella m
17	50	6.9	82	R818_BRUME	G9yfn9 bruceella m
18	49.5	6.8	127	LY6D_MOUSE	P3549 mus musculu
19	49	6.8	124	PRO2_PHYPO	P1832 physarum po
20	49	6.8	129	Y207_ABPPE	G9yfn2 aeropyrum p
21	49	6.8	130	CAL2_MOUSE	G99m3 mus musculu
22	48.5	6.7	110	AZLD_BACSU	O0793 bacillus su
23	48.5	6.7	126	YK50_ARCFU	P0387 archaeoglob
24	48.5	6.7	126	YK50_ARCFU	P0387 archaeoglob
25	48.5	6.7	132	BLE1_BACSP	P1304 bacillus sp
26	48	6.6	87	YAF2_BACLI	P3659 bacillus i
27	47.5	6.6	102	AMP1_MACIN	P80915 macadania i
28	47.5	6.6	65	HXB1_BRARE	O4236 brachydanio
29	47.5	6.6	83	IACA_PIG	P0099 sus scrofa
30	47.5	6.6	83	Y666_CHLTR	O8463 chlamydia t
31	47.5	6.6	84	Y710_CHLUP	G9z75 chlamydia p
32	47.5	6.6	92	Y710_CHLUP	G9z75 chlamydia p
33	47.5	6.6	92	Y710_CHLUP	G9z75 chlamydia p

34	47.5	6.6	97	IACS_PIG	P2641 sus scrofa
35	47.5	6.6	116	SMS_BOVIN	P2691 bos taurus
36	47.5	6.6	116	SMS_CANFA	P49670 canis famli
37	47.5	6.6	119	Y647_PASMU	P57862 pasteurella
38	47.5	6.6	129	YPS6_MYCTU	G50742 mycobacteri
39	47.5	6.6	129	YPS6_MYCTU	G50742 mycobacteri
40	47	6.5	120	MA5D_SCHUA	G9fn32 arabidopsis
41	47	6.5	121	RR22_GURTH	P1341 schistosoma
42	47	6.5	124	SSAP_SALTY	P74859 guillardiella
43	47	6.5	134	SPH1_SULSO	P4859 salmoneilla
44	47	6.5	134	SPH1_SULSO	P4859 salmoneilla
45	46.5	6.4	98	Y67_HPV11	P06322 cyclolobus
					P04020 human papil

## ALIGNMENTS

RESULT 1	ID	YIP3_SCHPO	STANDARD	PRT	94 AA
AC	G9t66	P79054			
DT	10-OCT-2003	(Rel. 42, Created)			
DT	10-OCT-2003	(Rel. 42, Last sequence update)			
DE	Hypothetical RING finger protein C343.03 in chromosome I.				
GN	SPAC343.03				
OS	Schizosaccharomyces pombe (Fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OC	Schizosaccharomycetes.				
ON	NCBI_TaxID=4896;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RX	MEDLINE=21849401; PubMed=11859360;				
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,				
RA	Gentles S., Goble A., Hamlin N., Harris D., Haldago J., Fraser A.,				
RA	Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,				
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,				
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,				
RA	Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Wolstencroft E., Aert R., Robben J., Gymnopoulos B.,				
RA	Welfjens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,				
RA	Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,				
RA	Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,				
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Jelaire V., Mottier S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Roche M., Galliard C., Tallada V.A., Garzon A., Thode G.,				
RA	Daga R.R., Cruzado L., Umenetz J., Sanchez M., del Rey F., Benito J.,				
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,				
RA	Cerrutli L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,				
RA	Shpakovskiy G.V., Useery D., Barrett B.G., Nurse P.,				
RA	"The genome sequence of Schizosaccharomyces pombe".				
RA	Nature 415:971-980(2002).				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=PR745;				
RX	MEDLINE=98162722; PubMed=9501991;				
RA	Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.,				
RT	"Identification of open reading frames in Schizosaccharomyces pombe				
RT	DNAS".				
RU	DNA Res. 4:363-369(1997).				
CC	-I- SIMILARITY: Contains 1 RING-type zinc finger.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AL109739; CAB52266.1; -  
 DR EMBL: AB01022; BA19316.1; ALT\_SEQ.  
 DR PIR: T3852; T3852.  
 DR GenDB\_Spombe; SPAC343.03; -  
 DR InterPro; IPR001841; ZnF\_ring.  
 DR SMART; SM00184; RING\_1.  
 DR PROSITE; PS00518; ZF\_RING\_1; FALSE\_NEG.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KM Hypothetical protein; Zinc\_finger.  
 FT ZN\_FING 35 78 RING-TYPE.  
 SQ SEQUENCE 94 AA; 10557 MW; C7770021339C3626 CRC64;

QY Query Match 7.7%; Score 56; DB 1; Length 94;  
 Best Local Similarity 42.9%; Pred. No. 28;  
 Matches 12; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

Db 24 GDSCPIYAGKPKMF-----IYNYVSDG 47  
 42 GDNCPIYWGKCKHIFAHCIQWLTSG 69

RESULT 2  
 HYBF\_KLEPN STANDARD; PRT; 113 AA.  
 ID HYBF\_KLEPN  
 AC Q9F0D4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable hydrogenase nickel incorporation protein hybf.  
 GN HYBF.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_Taxid=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=1157909;.  
 RA Fortlieau N.; Poirel L.; Nordmann P.;  
 RT "Plasmid-mediated and inducible cephalosporinase DNA-2 from Klebsiella  
 RT pneumoniae."  
 RL J. Antimicrob. Chemother. 47:207-210 (2001).  
 CC -1- FUNCTION: Probably plays a role in an hydrogenase nickel cofactor  
 CC insertion step.  
 CC -1- SIMILARITY: Belongs to the hybf/hybf family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AF295920; AAG36925.1; -  
 DR HAMAP: MF\_00213; -; 1.  
 DR InterPro; IPR000688; HyPA.  
 DR Pfam; PF0115; HyPA; 1.  
 DR Pfam; PF003620; HyPA; 1.  
 DR TIGRFAMs; TIGR00100; HyPA; 1.  
 DR PROSITE; PS01249; HyPA; 1.  
 KM Metal-binding; Nickel.  
 FT METAL 2 73  
 FT METAL 73 73 NICKEL (POTENTIAL).  
 FT METAL 76 76 NICKEL (POTENTIAL).  
 FT METAL 89 89 NICKEL (POTENTIAL).  
 FT METAL 92 92 NICKEL (POTENTIAL).  
 SQ SEQUENCE 113 AA; 12525 MW; 4D32362A9BF90516 CRC64;

QY Query Match 7.7%; Score 56; DB 1; Length 113;  
 Best Local Similarity 23.0%; Pred. No. 35;  
 Matches 23; Conservative 20; Mismatches 35; Indels 22; Gaps 5;

Db 44 VSDGLEDSSILEVDGPMK---NVEFKAKRGICVTHRA-----DFWSICTADMSLLE 96  
 26 VTDVWLEVGALADVEESALHFCFDIACRDVTAQGCCTHIDVTPAQAWCD-CSEAEIYQ 84

QY 97 QSHSPPSYLQCLSOCLROERGRTIPSGGITESKDMHFSYL 136  
 85 HAGCCP---HCGSERLR-----IIEGDLRKSL 110

RESULT 3  
 YD24\_PSEPK STANDARD; PRT; 124 AA.  
 ID YD24\_PSEPK  
 AC O88N88;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hypothetical UPF0102 protein PPI324.  
 GN PPI324.  
 OS Pseudomonas putida (strain KT2440).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_Taxid=160488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22423060; PubMed=12534463;  
 RA Nelson K.E.; Weinel C.; Paulsen I.T.; Dodson R.J.; Hilbert H.;  
 RA Martins dos Santos V.A.P.; Fouts D.E.; Gill S.R.; Pop M.; Holmes M.;  
 RA Brinkac L.; Beanan M.; DeBoy R.T.; Daugherty S.; Kolonay J.;  
 RA Madupu R.; Nelson W.; White O.; Peterson J.; Khouri H.; Hance I.;  
 RA Chris Lee P.; Holtzapfe E.; Scanlan D.; Tran K.; Kozdzek A.;  
 RA Ueberbach T.; Rizzo M.; Lee K.; Kosack D.; Moesti D.; Wedler H.;  
 RA Lauber J.; Stjepandic D.; Hehseil J.; Streitz M.; Heim S.;  
 RA Kiewitz C.; Eisen J.A.; Timmis K.N.; Duesterhoeft A.; Tuemmler B.;  
 RA Praser C.M.;  
 RT "Complete genome sequence and comparative analysis of the  
 RT metabolically versatile Pseudomonas putida KT2440."  
 RL Environ. Microbiol. 4:799-808 (2002).  
 CC -1- SIMILARITY: Belongs to the UPF0102 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AB016779; AAN66948.1; -  
 DR TIGR; PPI324; -  
 DR HAMAP: MF\_00046; -; 1.  
 DR InterPro; IPR003509; UPF0102.  
 DR Pfam; PF02021; UPF0102; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 124 AA; 13741 MW; 37A7471D485DB2P CRC64;

QY Query Match 7.7%; Score 56; DB 1; Length 124;  
 Best Local Similarity 23.3%; Pred. No. 40; Length 124;  
 Matches 21; Conservative 17; Mismatches 32; Indels 20; Gaps 4;

Db 29 YIAGKPKMFIONYVSDGLEDSSILEVD-----GPMKAVEFKAKRG 73  
 21 YIAGGGLQLLANRWKRGSEL-DLVMLDADTVFEVRYRIHAGFGALDISDKOKRRL 79

QY 74 L--CTVHREADPFWS--ICTADMSLLEQSH 99  
 80 VLAATLFLQKAHWGNYPCRFVVALQSH 109

```

RESULT 4
SOR_ARCFU STANDARD; PRT; 125 AA.
ID OSR_ARCFU
AC D29603;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Purative superoxide reductase (EC 1.15.1.2) (SOR).
GN AF0344.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
CX NCBI_taxonomy:2234;
[1]
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9399475;
RA Kleink H.-P., Claydon R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kirschum K.A., Dodson R.U., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Keziavaga A.R., Graham D.E., Kyprides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glocke A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utechtack T.,
RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Ruff C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter U.C.
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC CC
CC -1- FUNCTION: USES ELECTRONS FROM REDUCED NADP, BY WAY OF RUBREDOXIN
CC AND AN OXIDOREDUCTASE, TO CATALYZE THE REDUCTION OF SUPEROXIDE TO
CC HYDROGEN PEROXIDE. (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Reduced rubredoxin + superoxide + 2 H(+) =
CC rubredoxin + H(2)O(2).
CC -1- COFACTOR: Iron (By similarity).
CC -1- SIMILARITY: Belongs to the desulfoferrodoxin family.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@sib-sib.ch).
CC -----
CC CC EMBL: A0001081; AAB90892.1; -.
CC DR PIR: H69292; H69292.
CC DR HSPR; P82385; 1DQ1.
CC DR TIGR; AF0344; -.
CC DR InterPro: IPR002742; Desulfoferrodox.
CC DR Pfam: PF01880; Desulfoferrodox_1.
CC DR ProDom: PD006618; Desulfoferrodox_1.
CC DR TIGRFAMs: TIGR00332; neela_ferrous; 1.
CC KM Hypothetical protein; Oxidoreductase; Electron transport; Iron;
CC Complete proteome.
CC FT METAL 12 12 IRON (BY SIMILARITY).
CC FT METAL 14 14 IRON (BY SIMILARITY).
CC FT METAL 40 40 IRON (BY SIMILARITY).
CC FT METAL 46 46 IRON (BY SIMILARITY).
CC FT METAL 110 110 IRON (BY SIMILARITY).
CC FT METAL 113 113 IRON (BY SIMILARITY).
CC SQ SEQUENCE 125 AA; 13696 MW; D288A561SCDDDA7 CRC64;
Query Match 7.5%; Score 54.5; DB 1; Length 125;
Best Local Similarity 23.7%; P-vec. No. 59;
Matches 22; Conservative 9; Mismatches 29; Indels 33; Gaps 5;
0Y 5 GVDQTHSGPLNHHIRRMFM-----GDSPFYAGKPKMFIQNVVVSQGLQLESLSLEYVG 59
0b 35 GKRIHPNTHNHIAIMILVLPVQDSKRPVYVGAAR-----AAHG-----ASVDG 80

```

```

Dy 60 PAMKV-----EFKQKRG-----LCTV 78
Db 81 PNTSGVYTDPAVFAFKAKSGKUTAFSYCNTH 113

RESULT 5
CABP_SCHEMA
ID CABP_SCHEMA STANDARD; PRT; 69 AA.
AC P13566;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 13, Last annotation update)
DE Calcium-binding protein (CABP).
OS Schistosoma mansoni (Blood fluke).
CC Schistosoma matroidea; Platyhelminthes; Trematoda; Digenea; Strigeidida;
CC Schistosomatidae; Schistosomatidae; Schistosoma.
CX NBI_Ltaxid=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89219167; PubMed=2710168;
RA Ram D., Grossman Z., Markovics A., Avivi A., Ziv E., Lannner F.,
RA Schechter I.,
RT "rapid changes in the expression of a gene encoding a calcium-binding
RT protein in Schistosoma mansoni."
RL Mol. Biochem. Parasitol. 34:167-175(1989).
CC -1- DEVELOPMENTAL STAGE: This protein is expressed in the cercaria
CC but not in the sporocyst or adult worm.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04641; AAA29860.1; -.
DR EMBL: J04642; AAA29861.1; -.
DR PIR: A30792; A30792.
DR HSSP: P31950; 10LS.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 2.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00054; EFh; 2.
DR PROSITE: PS00018; EF_HAND; 2.
KM Calcium-binding; Repeat.
DT CA_BIND 15 26 EF-HAND 1 (POTENTIAL).
FT CA_BIND 51 62 EF-HAND 2 (POTENTIAL).
SQ SEQUENCE 69 AA; 7944 MW; DD2456636F8F876 CRC64;

Query Match 7.4%; Score 53.5; DB 1; Length 69;
Best local similarity 26.2%; Pred. No. 35;
Matches 17; Conservative 10; Mismatches 25; Indels 13; Gaps 1;

Qy 6 VQGTSGLEPLHHRMFGD-----SCPYLAKRKMFFIQQNVVSGQLEDS 52
Db 2 VNRTPAQLLKLHLPDRDSSGKISQELMFLHYVNCPEKKEQVEKIKQHDXDGGQLNTD 61
Qy 53 SLELV 57
Db 62 ELDDV 66

RESULT 6
SOR_PYRFU
ID SOR_PYRFU STANDARD; PRT; 124 AA.
AC P82385;
DT 15-OCT-2001 (Rel. 40, Created)
DT 15-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Superoxide reductase (EC 1.15.1.2) (SOR).
EN SORA OR PFI281.
```

OS Pyrococcus furiosus.  
 OC Archaea; Buryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RX MEDLINE=99445924; PubMed=10514376;  
 RA Jensen F.E., Jr., Verhagen M.F.J.M., Cui X., Adams M.W.W.;  
 RT "Anaerobic microbes: oxygen detoxification without superoxide  
 dismutase";  
 RL Science 286:306-309 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;  
 RT "The complete sequence of the Pyrococcus furiosus genome";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=20170875; PubMed=10704199;  
 RA Yeh A.P., Hu Y., Jensen F.E., Jr., Adams M.W.W., Rees D.C.;  
 RT "Structures of the superoxide reductase from Pyrococcus furiosus in  
 the oxidized and reduced states";  
 RL Biochemistry 39:2499-2508 (2000).  
 CC -1 FUNCTION: USES ELECTRONS FROM REDUCED NADP, BY WAY OF RUBREDOXIN  
 AND AN OXIDOREDUCTASE, TO CATALYZE THE REDUCTION OF SUPEROXIDE TO  
 HYDROGEN PEROXIDE.  
 CC -1 CATALYTIC ACTIVITY: Reduced rubredoxin + superoxide + 2 H(+) =  
 rubredoxin + H(2)O(2).  
 CC -1 COFACTOR: Iron.  
 CC -1 SUBUNIT: Homotetramer.  
 CC -1 SIMILARITY: Belongs to the desulfoferredoxin family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF156097; AF03229.1; -;  
 DR EMBL: AB010234; AB181405.1; -;  
 DR PIR: T44571; T44571.  
 DR PDB: 1DO6; 24-MAR-00.  
 DR PDB: 1DOJ; 10-MAY-00.  
 DR PDB: 1DOK; 10-MAY-00.  
 DR InterPro: IPR002742; Desulfoferredox.  
 DR Pfam: PF01880; Desulfoferredox.1.  
 DR ProDom: PD006618; Desulfoferredox.1.  
 DR TIGRFam: TIGR00332; neela\_ferrous.1.  
 KW Oxidoreductase; Electron transport; Iron; 3D-structure;  
 KW Complete proteome.  
 FT METAL 14 14 IRON.  
 FT METAL 16 16 IRON.  
 FT METAL 41 41 IRON.  
 FT METAL 47 47 IRON.  
 FT METAL 111 111 IRON.  
 FT METAL 114 114 IRON.  
 FT HELIX 2 5  
 FT STRAND 6 7  
 FT TURN 11 11  
 FT TURN 12 13  
 FT STRAND 14 14  
 FT TURN 16 16  
 FT TURN 19 19  
 FT STRAND 26 26  
 FT STRAND 28 34  
 FT STRAND 49 58  
 FT TURN 59 60  
 FT STRAND 65 72  
 FT STRAND 75 75  
 FT STRAND 78 78

FT TURN 79 80  
 FT STRAND 81 81  
 FT TURN 82 83  
 FT STRAND 88 89  
 FT STRAND 92 98  
 FT STRAND 103 111  
 FT TURN 112 114  
 FT STRAND 115 124  
 SQ SEQUENCE 124 AA; 14324 MW; EDD92C7E501C8048 CRC64;  
 Query Match 7.4%; Score 53.5; DB 1; Length 124;  
 Best Local Similarity 23.9%; Pred. No. 76;  
 Matches 22; Conservative 17; Mismatches 22; Indels 31; Gaps 6;  
 QY 5 GVDQTHSGLPDHPHIRMFMGDSCEYLAGKPMFPI---QNYVYSDGLEDSSILE-VDGP 60  
 DB 36 GKEIHPHTTEHNR-----YI---ELVLPGENFVYQVGRVETANGESVNGP 82  
 QY 61 AMKNV-----EFKAQKRG-----LCTVH 78  
 DB 83 NTSDVYTPPIAVFVLKTKKKGKLYALSYCNH 114  
 RESULT 7  
 YG73 METUA STANDARD; PRT; 129 AA.  
 ID YG73 METUA  
 AC Q59067;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein M1673.  
 GN M1673.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 CC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=868087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Petráň J.L., Nguyen D.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Rhee V.M., Gloeck A.,  
 RA Ullrich B., Kelley J.M., Peterson J.D., Sedow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hirst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RL Science 273:1058-1073 (1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U67607; AAB99695.1; -;  
 DR PIR: G64508; G64508.  
 DR TIGR: M1673; -;  
 KW Hypothetical protein; Complete proteome.  
 KM SEQUENCE 129 AA; 15418 MW; 06277ED71FD85C3E CRC64;  
 Query Match 7.2%; Score 52.5; DB 1; Length 129;  
 Best Local Similarity 20.5%; Pred. No. 1e+02;  
 Matches 17; Conservative 17; Mismatches 26; Indels 23; Gaps 3;  
 QY 34 PKGFYQNYVYSDGLEDSSILEVDGPMKRVKFAKRGICTVHRADPFWSICTADMS 93  
 DB 2 PKGFILFSKLTDDQIND-----ARKKLKVD-----FTYPRKLGELMSIIPDVD 48

QY 94 LLE-----QSHSPSLYU 106  
 Db 49 DIDNYLKPKEFLERKAKENDY 71

## RESULT 8

HYBF MORMO STANDARD; PRT; 113 AA.

AC OZHRI;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable hydrogenase nickel incorporation protein hybf.  
 GN HYBF.  
 OS Morganelle morganii (Proteus morganii).  
 SC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Morganelle.  
 CC NCBI\_TaxID=582;

RC STRAIN=GUI-1;  
 RA Polirel L., Guilbert M., Girlich D., Naas T., Nordmann P.;  
 RT "Cloning, sequence analyses, and distribution of ampc-ampR from  
 RT Morganelle morganii clinical isolates";  
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Probably plays a role in an hydrogenase nickel cofactor  
 insertion step.  
 CC -1- SIMILARITY: Belongs to the hybf/hybf family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF05067; AAC6580.1; -;  
 DR HAMAP; MF\_00213; -1;  
 DR InterPro: IPR000688; HyPA.  
 DR Pfam; PF01155; HyPA; 1.  
 DR ProDom; PD003620; HyPA; 1.  
 DR TIGRFAMs; TIGR00100; hyPa; 1.  
 DR PROSITE; PS01249; HyPA; 1.  
 KM Metal-binding; Nickel.  
 FT METAL 2 2  
 FT METAL 73 73 NICKEL (POTENTIAL).  
 FT METAL 76 76 NICKEL (POTENTIAL).  
 FT METAL 89 89 NICKEL (POTENTIAL).  
 FT METAL 92 92 NICKEL (POTENTIAL).  
 SQ SEQUENCE 113 AA; 12511 MW; 4D32362A9BF5DE16 CRC64;

Query Match 7.2%; Score 52; DB 1; Length 113;  
 Best Local Similarity 22.0%; Pred. No. 98;  
 Matches 22; Conservative 21; Mismatches 35; Indels 22; Gaps 5;

QY 44 VSDGQLEDSILEVGPAMK---NVEFPAQRGICVTHREA---DFNSICTADMSLLE 95  
 Db 26 VTDVWLEVGALADVESALHFDPAAGCDYVAGCTGTHIDVTPAQAWCWD-CSEAEIMQ 84  
 QY 97 QSHSPSLYUQCLSQKLRQERGITFGSGITSESKMDHFSFL 136  
 Db 85 HAGCCP---HCGSRRLR-----ISGGDLRVKSL 110

RESULT 9  
 SMS\_CHICK STANDARD; PRT; 116 AA.

AC P33094;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Somatostatin precursor [Contains: Somatostatin-28; Somatostatin-14].

GN SST.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 CC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Nata K., Kobayashi T., Karahashi K., Kato S., Yamamoto H.,  
 RA Yonekura H., Okamoto H.,  
 RL Submitted (JUN-1991) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Somatostatin inhibits the release of somatotropin.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to the somatostatin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X60191; CAA42747.1; -;  
 DR PIR; S20630; S20630.  
 DR InterPro: IPR004250; Somatostatin.  
 DR Pfam; PF03002; Somatostatin; 1.  
 KM Cleavage on pair of basic residues; Hormone; Signal.  
 FT SIGNAL 1 24  
 FT PROPEP 25 88 BY SIMILARITY.  
 FT PEPTIDE 89 116 SOMATOSTATIN-28.  
 FT PEPTIDE 103 116 SOMATOSTATIN-14.  
 FT DISULFID 105 116  
 SQ SEQUENCE 116 AA; 12675 MW; 8A5B9BD8A2918A CRC64;

Query Match 7.1%; Score 51.5; DB 1; Length 116;  
 Best Local Similarity 24.2%; Pred. No. 1.2e+02;  
 Matches 23; Conservative 14; Mismatches 45; Indels 13; Gaps 4;

QY 9 THSGLELHIRMFMGSCPYLACKPKM--FFICNYVDSQGLSDSLVD--GRAMK 63  
 Db 21 TVSAAPSDPRLRQFLQSLAAAGQELAKIFLAEILSEPSQTEALESEDSRGAED 80  
 QY 64 NVEFPAQR-----GLCTVHREA---DFNSICTA 90  
 Db 81 EVRLERLSANSNPALAPREKACKPFWKTFRS 115

RESULT 10  
 RPOL\_AERPE STANDARD; PRT; 102 AA.  
 ID RPOL\_AERPE  
 AC Q9YE21;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase subunit L (EC 2.7.7.6).  
 GN RPOL OR APE0443.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
 OC Desulfurococcaceae; Aeropyrum.

CC NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KL;  
 RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hailkawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Araki A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takemura M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,

RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1."  
 RL DNA Res. 6:83-101(1995).  
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC SIMILARITY: Belongs to the archaeobacteria RPO1 / eukaryotic RPB1/  
 CC RPB19 RNA polymerase subunit family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: A9000059; BA979405.1; -  
 CC PIR: A72739; X72739.  
 DR HANAP; NF\_002611; -1.  
 DR InterPro: IPR009025; RBP11-like\_RNapo.  
 DR InterPro: IPR008193; RNA\_pol\_L.  
 DR InterPro: IPR008194; RNA\_pol\_L\_sub.  
 DR Pfam: PF01193; RNA\_pol\_L\_1\_sub.  
 DR ProDom: PD004240; RNA\_pol\_L\_sub; 1.  
 DR PROSITE; PS01154; RNA\_POL\_L\_3KD; 1.  
 DR TransErase; Transcription; DNA-directed RNA polymerase;  
 KW Complete proteome.  
 SQ SEQUENCE 102 AA; 11930 MW; A87B560D92A15254 CRC64;  
 Query Match 7.0%; Score 51; DB 1; Length 102;  
 Best Local Similarity 27.8%; Pred. No. 1,1e+02;  
 Matches 22; Conservative 9; Mismatches 40; Indels 8; Gaps 1;  
 QY 3 VYGVDTQHSGLPHHTRKFMGDSCLYLAGKPPMFTIQYVVSQGLE-----DSSL 54  
 DB 21 VYGEDHTLGLTLEALRYVSNPKLAYEAVFWMEDIIQYVYEDVDVKEVLRASDYL 80  
 QY 55 LEVDGPAKMKVFEAKMG 73  
 DB 81 LEVIGDFRRRLALEERK 99  
 RESULT 11  
 FERR2 CAUCR STANDARD; PRT; 106 AA.  
 ID FERR2 CAUCR STANDARD; PRT; 106 AA.  
 AC P37058; 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ferredoxin, 2Fe-2S (FdiI).  
 GN Ferredoxin, 2Fe-2S OR C03524.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RA Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 RX MEDLINE=95270612; PubMed=7551304;  
 RA Wang S.P., Chen Y.P., Ely B.,  
 RT "A ferredoxin, designated Fdxp, stimulates p-hydroxybenzoate  
 RT hydroxylase activity in Caulobacter crescentus."  
 RL J. Bacteriol. 177:2908-2911(1995).  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gelin M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Emoliava M., White O.,  
 RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
 RT "Complete genome sequence of Caulobacter crescentus."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 CC -1- SIMILARITY: Belongs to the adenosine / putidaredoxin family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X51607; CA35950.1; -  
 CC EMBL: A806011; AA25466.1; -  
 CC PIR: S32573; S32573.  
 DR HSSP; P00259; IGFX.  
 DR TIGR; C03524; -  
 DR InterPro: IPR001055; Adenosine.  
 DR InterPro: IPR001041; Ferredoxin.  
 DR Pfam: PF00111; fer2; 1.  
 DR PRINTS; PR00355; ADRENOCOXIN.  
 DR PROSITE; PS00814; Adx; 1.  
 DR Electon transport; Metal-binding; Iron-sulfur; Iron; 2Fe-2S;  
 KW Complete proteome.  
 FT METAL 40 40 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 46 46 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 49 49 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 86 86 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 106 AA; 11366 MW; 59D096A72D993339 CRC64;  
 Query Match 7.0%; Score 51; DB 1; Length 106;  
 Best Local Similarity 29.6%; Pred. No. 1,2e+02;  
 Matches 24; Conservative 8; Mismatches 35; Indels 14; Gaps 5;  
 QY 42 VYVDSG--OLED--SLLEVDGPAKMKVE--FPAKRGK--TVHRADEFWSICTADMS 93  
 DB 6 YTHDGAECVIVKRGTLTWEGAVYCNVPGIDADCGACACATGVVDEMDLTKDGS 65  
 QY 94 LLEQS-----HSSPSLYLQC 108  
 DB 66 AMEESWLDFAENVFNSRLSC 86  
 RESULT 12  
 Y990 METUA STANDARD; PRT; 75 AA.  
 ID Y990 METUA STANDARD; PRT; 75 AA.  
 AC O58357; 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein M00990.  
 GN M00990.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RA MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,



RA Scott J.L., Geochagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Kienle H.-P., Fraser C.M., Smith H.O., Wiese C.R., Venter J.C.,  
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RA jannaschii".  
RL Science 273:1058-1073(1996).  
CC -1- SIMILARITY: Belongs to the UPF003 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U67541; AAB88992.1; -  
CC PIR: F64423; F64423.  
CC TIGR: M00990; -  
CC InterPro: IPR01455; UPF003.  
CC Pfam: PF01206; UPF003; 1.  
CC PROSITE: PS01148; UPF003; 1.  
CC Hypothetical protein: Complete proteome.  
CC DOMAIN 27 36 GLU-RICH;  
CC SEQUENCE 75 AA; 8465 MW; CF7009364C388539 CRC64;  
CC -----  
Query Match 7.0%; Score 50.5; DB 1; Length 75;  
Best Local Similarity 30.0%; Pred. No. 85;  
Matches 18; Conservative 7; Mismatches 24; Indels 11; Gaps 2;  
-----  
QY 24 GSCPTLAKPKPMFRTQNVVSDGLDSSLLIEDG--PMKQVEPKAGKGLCTVHRE 80  
DB 12 GDCIPVPLKTKKAL-----ELNIGCELEFVGVGYKPALENIKRFAENNGYVVLAE 63  
-----  
RESULT 13  
SMS HUMAN  
ID -SMS HUMAN STANDARD; PRT; 116 AA.  
AC P01166;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Somatostatin precursor (Contains: Somatostatin-28, Somatostatin-14).  
GN SST.  
OS Homo sapiens (Human), and  
OS Macaca fascicularis (Crah eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCB1\_TaxID=9606, 9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human.  
RX MEDLINE=84146798; PubMed=6142331;  
RX Shen L.-P., Rutter W.J.,  
RT "Sequence of the human somatostatin I gene";  
RL Science 224:168-171(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human.  
RX MEDLINE=83014931; PubMed=6126875;  
RX Shen L.-P., Pictet R.L., Rutter W.J.,  
RT "Human somatostatin I: sequence of the cDNA";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:4575-4579(1982).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human; TISSUE=Fetal brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RX Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RX Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schlier G.D.,  
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RX Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
RX Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.U., Ushin T.B., Toshiyuki S., Carlini P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kertman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman U.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.fascicularis;  
RX MEDLINE=88144503; PubMed=2894033;  
RX Travis G.H., Sutcliffe J.G.,  
RA "Phenol emulsion-enhanced DNA-driven subtractive cDNA cloning:  
RT isolation of low-abundance monkey cortex-specific mRNAs".  
RL Proc. Natl. Acad. Sci. U.S.A. 85:1696-1700(1988).  
CC -1- FUNCTION: Somatostatin inhibits the release of somatotropin.  
CC -1- SUBCELLULAR LOCATION: secreted.  
CC -1- PHARMACEUTICAL: Available under the name Sandostatin (Novartis);  
CC this is a synthetic cyclic analog known as octreotide or SMS  
CC 201-995. Used for the treatment of a variety of disorders  
CC including acromegaly and the symptomatic treatment of carcinoid  
CC tumors and vasoactive intestinal peptide tumors.  
CC -1- SIMILARITY: Belongs to the somatostatin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: J00306; AAB60566.1; -  
CC EMBL: M19318; AAB36908.1; -  
CC EMBL: BC032625; AAB32625.1; -  
CC PIR: A48968; A28968.  
CC PIR: A48614; R1HUS1.  
CC GeneW; HGNC:11329; SST.  
CC MIM: 182450; -  
CC GO: GO:0005615; C:extracellular space; TAS.  
CC GO: GO:0005180; F:peptide hormone; TAS.  
CC GO: GO:0007267; P:cell-cell signaling; TAS.  
CC GO: GO:0007586; P:digestion; TAS.  
CC GO: GO:0007186; P:G-protein coupled receptor protein signalin. .; TAS.  
CC GO: GO:0006288; P:induction of apoptosis by hormones; TAS.  
CC GO: GO:0008285; P:negative regulation of cell proliferation; TAS.  
CC GO: GO:0007268; P:response to nutrients; TAS.  
CC GO: GO:0007268; P:synaptic transmission; TAS.  
CC InterPro: IPR004250; Somatostatin.  
CC DR Cleavage on pair of basic residues; Hormone; Signal; Pharmaceutical.  
KW SIGNAL  
FT 1 24  
FT PROPEP 25 88  
FT PEPTIDE 89 116 SOMATOSTATIN-28.  
FT PEPTIDE 103 116 SOMATOSTATIN-14.  
FT DISULFD 105 116  
SQ SEQUENCE 116 AA; 12735 MW; AB49BB89DC9D8DA CRC64;  
-----  
Query Match 7.0%; Score 50.5; DB 1; Length 116;  
Best Local Similarity 23.5%; Pred. No. 1.5e+02;  
Matches 24; Conservative 13; Mismatches 34; Indels 31; Gaps 4;  
-----  
QY 11 GGLPIHRIKRFMGSCPYLAGKPKM--PFITQNVVSDGLDSSLLIEDG-- 54  
DB 23 TGAPSDRLRQFLQKSLAAAGKQELAKYFLAEILSEPNQTEADALEPDLQAPQDEM 82

```

OY      55-LEV-----DGPAMKNVEFFKQKRGGLCTVAREADPFWSLCTA 90
          |||         |||         |||         |||         |||
DB      83 RLEFLORSANSNPAMPAPREKKAGCK-----NFWMKFTFS 115

RESULT 14
SMS_SHEEP ID SMS_SHEEP STANDARD; PRT; 116 AA.
AC O46688;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Somatostatin precursor [Contains: Somatostatin-28; Somatostatin-14].
GN SST.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_taxid=9940;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ile de France;
RX MEDLINE=99094691; Pubmed=9880082;
RA Brunau G., Tillet Y.;
RT "Localization of the preprosomatostatin-mRNA by in situ hybridization
RT in the ewe hypothalamus";
RL Peptides 13:1749-1758(1998).
CC -1- FUNCTION: Somatostatin inhibits the release of somatotropin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the somatostatin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF031488; AAC094697.1; -.
DR EMBL; Y15267; CAAT5556.1; -.
DR InterPro; IPR004250; Somatostatin.
DR Pfam; PF03002; Somatostatin; 1.
KW Cleavage on pair of basic residues; Hormone; Signal.
FT SIGNAL 1 24 BY SIMILARITY.
FT PROPEP 25 88 BY SIMILARITY.
FT PEPTIDE 89 116 SOMATOSTATIN-28.
FT PEPTIDE 103 116 SOMATOSTATIN-14.
FT DISULFID 105 116 BY SIMILARITY.
SQ SEQUENCE 116 AA; 12689 MW; C1SF7EJ3IA37I8DE CRC64;

Query March 7.0%; Score 50.5; DB 1; Length 116;
Best Local Similarity 23.5%; Pred. No. 1.5e+02;
Matches 24; Conservative 13; Mismatches 34; Indels 31; Gaps 4.

OY 11 SGPLPHIRMFPGDCPYTLACKPKM--FFLQNVVVSDGQLQSLSL----- 54
DB 23 TGAPSDPRLFQFKLAALAAKQKLAKYFLAEILSEPNQENDALEPEDLSQALEDQM 82
OY 55 -LEV-----DGPAMKNVEFFKQKRGGLCTVAREADPFWSLCTA 90
          |||         |||         |||         |||         |||
DB 83 RLEFLORSANSNPAMPAPREKKAGCK-----NFWMKFTFS 115

RESULT 15
GLUC_CARAU STANDARD; PRT; 121 AA.
AC P79695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glucagon precursor [Contains: Glucocentin-related polypeptide (GRPP);

```

```

01 Glucagon: Glucagon-like peptide).
02
03 Carassius auratus (goldfish).
04
05 Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;
06 Actinopterygii, Neopterygii, Teleostei, Ostariophysi, Cypriniformes;
07 Cyprinidae; Carassius.
08 NCBI_TaxID=7957;
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
```

```
Search completed: September 9, 2004, 06:23:11
Job time : 12 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:22:13 (without alignments)

1137.526 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_205\_341

Sequence: 1 OSVGVDDTHSGPLPHIR.....CTPGSGITSGKDWHPFSLG 137

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 279462

Minimum DB seq length: 0

Maximum DB seq length: 137

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	9.0	103	Q7V6Y6	Q7V6Y6 prochlorococcus
2	62.5	8.6	89	Q93J15	Q93J15 streptomyces
3	60.5	8.3	100	Q8YPP6	Q8YPP6 anabaena sp
4	60.5	8.3	100	Q8N138	Q8N138 homo sapien
5	59.5	8.2	102	Q8HXA1	Q8HXA1 macaca fasc
6	59.5	8.2	129	Q9LFR4	Q9LFR4 arabidopsis
7	59.5	8.2	131	Q93J15	Q93J15 arabidopsis
8	58.5	8.1	83	Q99DD6	Q99DD6 streptomyces
9	58	8.0	62	Q8VAX8	Q8VAX8 white spot
10	57.5	7.9	81	Q8ZUS6	Q8ZUS6 enterobacter
11	57	7.9	99	Q858E4	Q858E4 enterobacter
12	57	7.9	109	Q8ZX21	Q8ZX21 pyrobaculum
13	57	7.9	120	Q9LJ14	Q9LJ14 arabidopsis
14	56.5	7.8	91	Q8W4S9	Q8W4S9 arabidopsis
15	56.5	7.8	124	Q8PR13	Q8PR13 xanthomonas
16	56	7.7	100	Q27020	Q27020 methanobact

17	56	7.7	115	3	Q9P8V1	Q9P8V1 ashbya goss
18	56	7.7	124	16	Q88N88	Q88N88 pseudomonas
19	56	7.7	129	16	Q8K266	Q8K266 chlorobium
20	56	7.7	132	2	Q9F7M0	Q9F7M0 mycobacteri
21	56	7.7	132	2	Q9KH55	Q9KH55 mycobacteri
22	55.5	7.7	130	16	Q8ABV3	Q8ABV3 bacteroides
23	55.5	7.7	131	10	Q8L6S8	Q8L6S8 arabidopsis
24	55.5	7.7	134	12	Q65799	Q65799 mycobact
25	55	7.6	78	12	Q7U750	Q7U750 synchococc
26	54.5	7.5	91	12	Q9IC09	Q9IC09 oropouche v
27	54.5	7.5	91	12	Q9ICD9	Q9ICD9 oropouche v
28	54.5	7.5	91	12	Q9J951	Q9J951 oropouche v
29	54.5	7.5	91	12	Q9J973	Q9J973 oropouche v
30	54.5	7.5	91	12	Q9ICD8	Q9ICD8 oropouche v
31	54.5	7.5	101	16	Q9P8S5	Q9P8S5 xylella fas
32	54.5	7.5	117	10	Q40553	Q40553 nicotiana t
33	54.5	7.5	129	17	Q970N8	Q970N8 sulfolobus
34	54	7.4	98	2	Q9JPF9	Q9JPF9 neisseria m
35	54	7.4	98	2	Q9JPD6	Q9JPD6 neisseria m
36	54	7.4	114	10	Q8GZ27	Q8GZ27 oryza sativ
37	54	7.4	120	4	Q8H7M6	Q8H7M6 homo sapien
38	54	7.4	125	16	Q8F520	Q8F520 leprospira
39	54	7.4	136	17	Q8ZX25	Q8ZX25 pyrobaculum
40	53.5	7.4	91	12	Q9J956	Q9J956 oropouche v
41	53.5	7.4	91	12	Q9ICE3	Q9ICE3 oropouche v
42	53.5	7.4	125	16	Q890N6	Q890N6 clostridium
43	53.5	7.4	137	2	Q93RR4	Q93RR4 rhodobacter
44	53	7.3	65	2	Q92G19	Q92G19 chlamydia t
45	53	7.3	111	4	Q9BRE8	Q9BRE8 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q7V6Y6 PRELIMINARY; PRT; 103 AA.  
AC Q7V6Y6;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Possible tRNA synthetases class I (E and Q).  
GN PMT0991.  
OS Prochlorococcus marinus (strain MIT 9313).  
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=74547;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2282698; PubMed=12917642;  
RA Kocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
RT niche differentiation."  
RT Nature 424:1042-1047(2003).  
RL EMBL; BX572098; CAE21166.1;  
SQ AMINOACYL-tRNA synthetase; Complete proteome.  
SEQUENCE 103 AA; 12135 MW; 6B06F84F2BCBFD0 CRC64;

Query Match 9.0%; Score 65; DB 16; Length 103;  
Best Local Similarity 31.7%; Pred. No. 16;  
Matches 20; Conservative 6; Mismatches 29; Indels 8; Gaps 2;

QY 77 VHRADFF--WSCTADMSILLESSSPELYQCUSQKURORGRITPGSGITSGKDWHPF 133  
DB 42 VFTITTFPFWIMSLAGIYLLSKSHRIQKQILHAKROMG-----ASTLLDSEHI 96  
QY 134 SSL 136  
DB 97 SKL 99

## RESULT 2

093J15 PRELIMINARY; PRT; 89 AA.  
 AC 093J15;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein SC03993.  
 DE SC03993 OR SCBAC2583.30C.  
 GN SC03993 OR SCBAC2583.30C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OC NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Laake L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL339118; CAC44717.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 89 AA; 10063 MW; 02FB9B00B1068C2 CRC64;

Query Match 8.4%; Score 62.5; DB 16; Length 89;  
 Best Local Similarity 36.4%; Pred. No. 25;  
 Matches 12; Conservative 9; Mismatches 7; Indels 5; Gaps 1;

QY 4 YGVDDTHSGLPPLHH-----IRRMFGDSCPYLA 31  
 DB 52 YVVENHTEGCPVHSARMPVDMKFEVGNACCVYA 84

## RESULT 3

08YPP6 PRELIMINARY; PRT; 87 AA.  
 AC 08YPP6;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein Asl4146.  
 GN ASL4146.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OC NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Matsumoto M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003595; BAB75845.1; -;  
 DR PIR; AC2324; AC2324.  
 DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR003115; PatBc.  
 DR Pfam; PF02195; PatBc; 1.

DR SMART; SMO0470; PatB; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 87 AA; 9918 MW; 72567AB41142EC3 CRC64;

Query Match 8.3%; Score 60.5; DB 16; Length 87;  
 Best Local Similarity 29.9%; Pred. No. 42;  
 Matches 20; Conservative 8; Mismatches 24; Indels 15; Gaps 2;

QY 13 LPLHIRMFGDSCPYLAGKPKMFYQNYVSDGQLEDSLSLEVDGPMKVFRAQR 72  
 DB 7 IPLNQIRSLPRGNDPF-----KQALMESTIALGQGFPIDVLEVVG-----QY 51

QY 73 GLCTVHR 79  
 DB 52 GFGGCHR 58

## RESULT 4

08N338 PRELIMINARY; PRT; 100 AA.  
 AC 08N338;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon, and Kidney;  
 RA Strausberg R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC028282; AAH28282.1; -;  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 100 AA; 11421 MW; B98D4B924B38F70 CRC64;

Query Match 8.3%; Score 60.5; DB 4; Length 100;  
 Best Local Similarity 26.0%; Pred. No. 50;  
 Matches 26; Conservative 14; Mismatches 27; Indels 33; Gaps 5;

QY 12 GLPLHIRMFGDSCPYLAGKPKMFYQNYVSDGQLEDSLSLEVDGP--AMKNVEFK 68  
 DB 17 GLPLHH-----RMFAIQ-----TEGLRATVQLKEKFPATIAIKVTDIS 55

QY 69 AQRKGLCTVHREADPFWSLCTADVSLLEQSHSSPSLYIQ 108  
 DB 56 GTRK-----RNQRDAM--IADYLSPTLTFTWLHRC 86

## RESULT 5

08HXEL PRELIMINARY; PRT; 102 AA.  
 AC 08HXEL;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OC NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;  
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;

```

FX MEDLINE=11458551; PubMed=11574149;
RA Oosada N., Hida M., Kusuda Y., Tanuma R., Ieki K., Hirata M., Suto Y.,
RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.,
RA "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
RT chromosomes."
RL Gene 275:31-37(2001).
DR EMBL; AB093649; BAC21623.1; -.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0003918; P:DNA topoisomerase type II activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006265; P:DNA topological change; IEA.
DR InterPro; IPR001241; DNA_topoisom.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW HYPOTHEICAL protein.
SQ SEQUENCE 102 AA; 10763 MW; DED69A2FE605794 CRC64;

Query Match 8.2%; Score 59.5; DB 6; Length 102;
Best Local Similarity 28.0%; Pred. No. 66;
Matches 28; Conservative 15; Mismatches 46; Indels 11; Gaps 5;

QY 44 VSDGQ-EDSSLLEV---DGMAMVNFKQKRGLCVHREADFFMSLCTADMSILFQS 98
Db 5 VTSDDLNEGNSALTYPDQTEAFSVMETASRSRGRLAT---QAMCTAGDHSILPLG 60
QY 99 HSSPSLYLQCLSQKLRQERGT-IPGSGITES-KMHPFSL 136
Db 61 SHAPSIQLACGPDLANQHPIPTGLGSPSIDRGAKDLHOSHL 100

RESULT 6
Q9LF84 PRELIMINARY; PRT; 129 AA.
AC Q9LF84;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Squamosa promoter binding protein-like 3.
GN Sp33.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryote; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Columbia; TISSUE=flower;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hehmann S., Nettushin K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene Sp33: a
RL novel gene involved in the floral transition."
RL Plant J. 12:367-377(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Columbia; TISSUE=flower;
RA Cardon G.H., Huijser P.;
RT "The SBP-box gene, family of plant DNA-binding proteins."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ242959; CAB94233.1; -.
DR PIR; H84749; H84749.
DR TRANSFAC; T05182; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR InterPro; IPR004333; SBP.
DR Pfam; PF03110; SBP; 1.
SQ SEQUENCE 129 AA; 15086 MW; 03BEFF6440A37574 CRC64;

Query Match 8.2%; Score 59.5; DB 10; Length 129;
Best Local Similarity 24.0%; Pred. No. 88;
Matches 23; Conservative 13; Mismatches 41; Indels 19; Gaps 3;

QY 32 GKRMFFIQNYVVSQGLDSSLLEVDGPMKNVFFKA-QRRGICTVHREADFFMSLCTA 90

```

[illegible]

DR EMBL; Y093427; CAAT0578.1; -  
 DR EMBL; U78721; AAC69133.2; -  
 DR EMBL; AJ011633; CAB56585.1; -  
 DR EMBL; AJ011627; CAB56579.1; -  
 DR EMBL; AK118179; BAC42802.1; -  
 DR EMBL; BT005443; AAO63863.1; -  
 DR PIR; T52597; T52597.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR004335; SBP.  
 DR Pfam; PF03110; SBP; 1.  
 SQ SEQUENCE 131 AA; 15304 MW; F3B4C7D8F6537501 CRC64;

Query Match 8.2%; Score 59.5; DB 10; Length 131;  
 Best Local Similarity 24.0%; Pred. No. 90;  
 Matches 23; Conservative 13; Mismatches 41; Indels 19; Gaps 3;

QY 32 GKPRMFIQNYVVSQGLSDSLVDGFMKNVFKK-OKKGLCTVREADFWSLCTA 90  
 Db 10 GKSLRLSEEESEETEDTFFEEELKXKKGKATSSSGVCQVES-----CTA 61  
 QY 91 DMSLLBQSH-----SSPSLYQCLSSOKLRQE 116  
 Db 62 DMSKAKOYHKKHVKCOFHAKAPHVRIISGLHQRFCQD 97

RESULT 8  
 Q990D6 PRELIMINARY; PRT; 83 AA.

AC Q990D6 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein SCPI.282.  
 GN SCPI.282 AND SCPI.71C.  
 OS Streptomyces coelicolor.  
 CC Plasmid SCPI.  
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1902;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL590464; CAC36808.1; -  
 DR EMBL; AL590463; CAC36593.1; -  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR Hypothetical protein; Plasmid; Complete proteome.  
 KW SEQUENCE 83 AA; 9522 MW; 6957D1E43CEC28B6 CRC64;

Query Match 8.1%; Score 58.5; DB 16; Length 83;  
 Best Local Similarity 36.5%; Pred. No. 68;  
 Matches 19; Conservative 4; Mismatches 20; Indels 9; Gaps 2;

QY 44 VSDQGLSDSLVDGPA--MKNVFKKQKGLCTVRE-----ADFFWS 86  
 Db 11 IODGPASDSSDTEESQPARLVKPEFGAPSRQAVTVLRLTETPEPSTGFWHS 62

RESULT 9  
 Q8VAX8

ID Q8VAX8 PRELIMINARY; PRT; 62 AA.  
 AC Q8VAX8;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Wsv241 (WSSV237).  
 OS White spot syndrome virus (WSSV).  
 CC Viruses; dsDNA viruses, no RNA stage; Nimaviridae.  
 OX NCBI\_TaxID=92652;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21548311; PubMed=11689662;  
 RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
 RT "Complete genome sequence of the shrimp white spot bacilliform  
 virus."  
 RL J. Virol. 75:11811-11820(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Taiwan;  
 RX MEDLINE=20517548; PubMed=11062040;  
 RA Tsai W.-F., Yu H.-T., Zheng H.-F., Leu J.-H., Chou C.-M., Huang C.-J.,  
 RA Wang C.-H., Lin J.-Y., Kou G.-H., Lo C.-F.;  
 RT "Identification and characterization of a shrimp white spot syndrome  
 RT virus (WSSV) gene that encodes a novel chimeric polypeptide of  
 RT cellular-type thymidine kinase and thymidylate kinase."  
 RL Virology 277:100-110(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Taiwan;  
 RX MEDLINE=21844071; PubMed=11853398;  
 RA Chen L.-L., Leu J.-H., Huang C.-J., Chou C.-M., Chen S.-M., Wang C.-H.,  
 RA Lo C.-F., Kou G.-H.;  
 RT "Identification of a nucleocapsid protein (VP3) gene of shrimp white  
 RT spot syndrome virus and characterization of the motif important for  
 RT targeting VP3 to the nuclei of transfected insect cells."  
 RL Virology 293:44-53(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Taiwan;  
 RA Lo C.-F., Kou G.-H.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF332093; AAL35245.1; -  
 DR EMBL; AF440570; AAL89165.1; -  
 SQ SEQUENCE 62 AA; 6698 MW; 9BBCDD06C5367855 CRC64;

Query Match 8.0%; Score 58; DB 12; Length 62;  
 Best Local Similarity 31.8%; Pred. No. 55;  
 Matches 21; Conservative 5; Mismatches 18; Indels 22; Gaps 4;

QY 69 AQRGLCTVREADFWSLCTADMSLLBQSHSPSLYLQCLSS-----OKLRQERGIT-PG 122  
 Db 8 ADRGGLTKEHS-----WHLVLADIS-----RECCSLFTTAKTKOSGTTSPG 51

QY 123 SGITES 128  
 Db 52 SVDDPS 57

RESULT 10  
 Q8ZUS6 PRELIMINARY; PRT; 81 AA.  
 AC Q8ZUS6;  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical protein PAE2633.  
 GN PAE2633.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;

```
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-U., Stetter K.O., Simon M.I.,
  Miller U.H.;
RT "genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
  aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009887; AAL64330.1; -.
DR InterPro; IPR001455; UPF0033.
DR Pfam; PF01206; UPF0033; 1.
DR PROSITE; PS01148; UPF0033; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 81 AA; 8820 MW; 2AF84D9D613061B6 CRC64;

Query Match
Best Local Similarity 31.1%; Pred. No. 86; Length 81;
Matches 19; Conservative 14; Mismatches 17; Indels 11; Gaps 3;

QY 24 GDSCEYLAKPKPMFICNYVSDGQLEDSLSLEV--DGPAKQVEFKAKRGKCTVHRE 80
Db 18 GYVCPY---PQWYTSQ---ALSKLPKGSVLKVIIDNPSEINIKSAVQKAKKSLQVE 69
QY 81 A 81
Db 70 A 70

RESULT 11
Q858B4 PRELIMINARY; PRT; 99 AA.
AC Q858B4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 31.
OS Enterobacteria phage epsilon15.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC P22-like viruses.
OX NCBI_TaxID=215158;
RN [1]
RP SEQUENCE FROM N.A.
RA Kropinski A.M., Billington S.J., Patrick A.N., Butts B.D.,
  Kovalyova I., McConnell M.R.;
RT "The Genomic Sequence of the Group E1 Salmonella enterica Converting
  Racteriophage, Epsilon 15."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY150271; AA06094.1; -.
SQ SEQUENCE 99 AA; 11244 MW; 4212FB0ADD0C642C CRC64;

Query Match
Best Local Similarity 30.6%; Pred. No. 1.2e+02; Length 99;
Matches 26; Conservative 14; Mismatches 39; Indels 6; Gaps 4;

QY 41 NYVSDGQLEDSLSLEVDPAMKQVEFKAKRGKCTVHREADFMSLCTADMSLLEQSHS 100
Db 13 NVV---SQRQAKKVAIVT-PAAMEVEKAAVERESKGYRIARLMLLC-MDVAVGEVERA 67
QY 101 SPSLYL-QCLSGKLRQERGTIPGSG 124
Db 68 RAIIRDQCISRGNGIRGEYSIG 92

RESULT 12
Q8ZX21 PRELIMINARY; PRT; 109 AA.
AC Q8ZX21;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE Hypothetical protein PAE1511.
GN PAE1511.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-U., Stetter K.O., Simon M.I.,
  Miller U.H.;
RT "genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
  aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009824; AAL63528.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 109 AA; 11984 MW; 603823462FDE0272 CRC64;

Query Match
Best Local Similarity 26.4%; Pred. No. 1.4e+02; Length 109;
Matches 19; Conservative 15; Mismatches 22; Indels 16; Gaps 4;

QY 8 QHNSGLPLHTRRMFMDSCPYLAGKPKMFIQNYVSDGQLEDSLSLEVDPAMKQVEF 67
Db 27 ELHNKPPDRFRLRATGDA-----FIEINISQSEID--WDEVYGAEGG-PF 71
QY 68 KAQRKGLCTVHR 79
Db 72 RPQ-RGLCAVNG 82

RESULT 13
Q9LIJ4 PRELIMINARY; PRT; 120 AA.
AC Q9LIJ4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Genomic DNA, chromosome 3, BAC clone:FBN14.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Kropinski A.M., Billington S.J., Patrick A.N., Butts B.D.,
  Kovalyova I., McConnell M.R.;
RT "Structural Analysis of Arabidopsis thaliana Chromosome 3. III."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001301; BAB01896.1; -.
DR InterPro; IPR007286; EAP30.
DR Pfam; PF04157; EAP30; 1.
SQ SEQUENCE 120 AA; 13629 MW; 7E7A0F11B3A53EF CRC64;

Query Match
Best Local Similarity 29.0%; Pred. No. 1.6e+02; Length 120;
Matches 18; Conservative 11; Mismatches 25; Indels 8; Gaps 2;

QY 60 PAMKQVEFKAKRGKCTVHREADFMSLCT-ADMSLLEQSHSPSLYQLCSQKLRQER 117
Db 9 PLASIKGFMAELTGI-----GDPEYIEIGVQITIEVCWLTFRSHNGKSLISQELCNHLRQR 62
QY 118 GT 119
Db 63 KT 64

RESULT 14
Q8W459 PRELIMINARY; PRT; 91 AA.
AC Q8W459;
```

DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Katlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Barn J., Garinaci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Saitou M., Seki M., Shimizu P., Yamada K., Shinzaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF062831; AAL32809.1; -.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 91 AA; 10166 MW; 23B136FE2CDD3AF1 CRC64;  
 Query Match 7.8%; Score 56.5; DB 10; Length 91;  
 Best Local Similarity 33.3%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 4; Mismatches 17; Indels 15; Gaps 2;  
 QY 98 SHSSPSLYLOCLS-----QKLRQERG-----TFGSGITSKDMHPSL 136  
 Db 22 SNSSSLFLTCLSSHYLTITKIRDRRGVEKSTGRATLTGFAEVNDVTMFPGGL 75  
 RESULT 15  
 Q8PR13 PRELIMINARY; PRT; 124 AA.  
 AC Q8PR13;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein XACB0049.  
 GN XACB0049.  
 OS Xanthomonas axonopodis (pv. citri).  
 OG Plasmid pXAC64.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 NCBI\_TaxID=92829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Ouagga R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
 RA Camarotte G., Camavan F., Cardozo J., Chambergo F., Chapina L.P.,  
 RA Cicarelli R.M.B., Coutinho L.V., Cursino-Santos J.R., El-Dorry H.,  
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Secubal J.C., Kitajima J.P.;  
 RL "Comparison of the genomes of two Xanthomonas pathogens with differing  
 RT host specificities";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AE008925; AAM59295.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 KW Hypothetical protein; Plasmid; Complete proteome.  
 SQ SEQUENCE 124 AA; 13581 MW; FB8563147CB69B82 CRC64;

Query Match 7.8%; Score 56.5; DB 16; Length 124;  
 Best Local Similarity 22.3%; Pred. No. 1.9e+02;  
 Matches 27; Conservative 11; Mismatches 46; Indels 37; Gaps 5;  
 QY 20 RMFMGDS-C-PYLAKPKMFFIQNTYVSD-----GQLEDS- 53  
 Db 11 RYFGSGGLPGLKRPSCFQAFVLDGRGWSVGVLOGLGEGSDHRRIGQAARESGS 70  
 QY 54 ---LIEVDGPMKQVFEFAQKRGCTVHREADFFWSL-CTADMSLEQSHSSPSLYLOCL 109  
 Db 71 VDGIYELINHSWQISROGDVHGCGFAHRSKFLWGLTMPIDWIFK-----LKRCL 123  
 QY 110 S 110  
 Db 124 A 124

Search completed: September 9, 2004, 06:25:25  
 Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 9, 2004, 06:22:03 ; Search time 53 Seconds  
(without alignments)  
730.359 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_205\_341

Perfect score: 725  
Sequence: 1 QSVYVDQTHSGPLPHIR.....GTFPGSGTSEKXMHFSLG 137

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 988826

Minimum DB seq length: 0  
Maximum DB seq length: 137

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1890s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	391	53.9	76	AA15359	AA15359 Peptide #
2	391	53.9	76	AB834356	AB834356 Peptide #
3	391	53.9	76	AA27837	AA27837 Peptide #
4	391	53.9	76	AB829199	AB829199 Peptide #
5	391	53.9	76	AB819775	AB819775 Peptide #
6	391	53.9	76	AA67546	AA67546 Human bon
7	391	53.9	76	AA655153	AA655153 Human bon
8	391	53.9	76	AB849189	AB849189 Human liv
9	391	53.9	76	AA03118	AA03118 Peptide #
10	391	53.9	76	AB837140	AB837140 Human pep
11	391	53.9	39	AA814254	AA814254 Partial m
12	61.5	8.5	101	AD808128	AD808128 Novel pro
13	60.5	8.3	80	AAU14746	AAU14746 Novel bon
14	60	8.3	72	AAU40343	AAU40343 Propionib
15	60	8.3	72	ABM36862	ABM36862 Propionib
16	60	8.3	92	AA833250	AA833250 Eucalyptu
17	59.5	8.2	124	AB828010	AB828010 Novel hum
18	59.5	8.2	129	AA836780	AA836780 Arabidops
19	59.5	8.2	131	AA770954	AA770954 Arabidops
20	59.5	8.2	131	AA836779	AA836779 Arabidops
21	59.5	8.2	131	AA830387	AA830387 Arabidops
22	59	8.1	117	AA811793	AA811793 Histo-blo
23	59	8.1	117	AA857017	AA857017 Human O t
24	58.5	8.1	94	AD894032	AD894032 E. faeciu
25	58.5	8.1	115	AB870778	AB870778 Human ext

26	58	8.0	71	3	AA60552	AA60552 Arabidops
27	58	8.0	72	3	AA856185	AA856185 Arabidops
28	58	8.0	99	4	AB809354	AB809354 Novel hum
29	57.5	7.9	64	4	AA890887	AA890887 Human imm
30	57	7.9	114	5	AB840423	AB840423 Human sec
31	57	7.9	115	3	AA832611	AA832611 Eucalyptu
32	57	7.9	115	3	AA833172	AA833172 Eucalyptu
33	57	7.9	118	3	AA833753	AA833753 Human sec
34	56.5	7.8	78	3	AA861267	AA861267 Arabidops
35	56.5	7.8	105	3	AA861256	AA861256 Arabidops
36	56.5	7.8	112	3	AA861255	AA861255 Arabidops
37	56.5	7.8	112	3	AA802957	AA802957 Human sec
38	56.5	7.8	122	4	AA884678	AA884678 Human imm
39	56.5	7.8	135	7	AD831087	AD831087 Human dia
40	56	7.7	77	4	AA864345	AA864345 Statmin
41	56	7.7	94	2	AA860215	AA860215 Human end
42	56	7.7	113	2	AA813504	AA813504 Thrombosp
43	55.5	7.7	68	6	AB819123	AB819123 Parthogen
44	55.5	7.7	129	3	AA823107	AA823107 Arabidops
45	55.5	7.7	131	3	AA823106	AA823106 Arabidops

ALIGNMENTS

RESULT 1  
AA15359  
ID AA15359 standard; protein; 76 AA.  
XX  
AC AA15359;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Peptide #1793 encoded by probe for measuring cervical gene expression.  
XX  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
XX  
KW cervical cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157278-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JUN-2001; 2001MO-US000670.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX  
PR 26-MAY-2000; 2000US-0207456P.  
XX  
PR 03-JUN-2000; 2000US-00608408.  
XX  
PR 03-AUG-2000; 2000US-00632366.  
XX  
PR 21-SEP-2000; 2000US-0234687P.  
XX  
PR 27-SEP-2000; 2000US-0236358P.  
XX  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488901/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.  
XX  
PS Claim 27; SEQ ID NO 20185; 487bp; English.  
XX  
CC The present invention relates to human single exon nucleic acid probes  
CC (SENs: see A110068-A118459). The present sequence is a peptide encoded  
CC by one such probe. The SENs are derived from human Hela cells. The SENs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 76 AA;

Query Match 53.9%; Score 391; DB 4; Length 76;

Best Local Similarity 98.7%; Pred. No. 1.8e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 41 NYVVSQGLDSSSLEVDGPAMKNVEFFKQKRGCTVHREADFWSLCTADMSLLEQSHS 100  
DB 1 NYVVSQGLDSSSLEVDGPAMKNVEFFKQKRGCTVHREADFWSLCTADMSLLEQSHS 60  
QY 101 SPSTLYQCLSQKLRQE 116  
DB 61 SPSTLYQCLSQKLRQE 76

RESULT 2

ABB34356 ID ABB34356 standard; peptide; 76 AA.

AC ABB34356;

DT 04-FEB-2002 (first entry)

DE Peptide #1862 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US000669.

PF 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

DR Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.

PT Claim 27; SEQ ID NO 26391; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human fetal liver. The

XX present sequence is a peptide encoded by a single exon nucleic acid probe

XX of the invention. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 76 AA;

DB 1 NYVVSQGLDSSSLEVDGPAMKNVEFFKQKRGCTVHREADFWSLCTADMSLLEQSHS 60

QY 101 SPSTLYQCLSQKLRQE 116

DB 61 SPSTLYQCLSQKLRQE 76

RESULT 3

AAM27837 ID AAM27837 standard; protein; 76 AA.

AC AAM27837;

DT 17-OCT-2001 (first entry)

DE Peptide #1874 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

DR Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

PT Claim 27; SEQ ID NO 28106; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

XX see A131315-A131546). The present sequence is a peptide encoded by one

XX CC such probe. The probes are useful for producing a microarray for

XX CC predicting, measuring and displaying gene expression in samples derived

XX CC from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders

XX Sequence 76 AA;

Query Match 53.9%; Score 391; DB 4; Length 76;

Best Local Similarity 98.7%; Pred. No. 1.8e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 41 NYVVSQGLDSSSLEVDGPAMKNVEFFKQKRGCTVHREADFWSLCTADMSLLEQSHS 100

DB 1 NYVVSQGLDSSSLEVDGPAMKNVEFFKQKRGCTVHREADFWSLCTADMSLLEQSHS 60

QY 101 SPSTLYQCLSQKLRQE 116

DB 61 SPSTLYQCLSQKLRQE 76

RESULT 4

ABB29199 ID ABB29199 standard; peptide; 76 AA.

```

XX AC ABB29199;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #1850 encoded by breast cell single exon nucleic acid probe.
XX KM Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-496933/54.
XX PT New spatially-addressable set of single exon nucleic acid probes, useful
XX PT for measuring gene expression in sample derived from human breast,
XX PT comprises number of single exon nucleic acid probes.
XX PS Claim 27; SEQ ID NO 12157; 327bp + Sequence Listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting the
XX CC probes with a collection of detectably labeled nucleic acids derived
XX CC from mRNA of human breast, and then measuring the label bound to each
XX CC probe of the microarray. The probes are useful for verifying the
XX CC expression of regions of genomic DNA predicted to encode proteins. They
XX CC are useful for gene discovery, and for determining predisposition and/or
XX CC prognosing breast disease. Gene expression analysis is useful for
XX CC assessing the toxicity of chemical agents on cells. The microarray of
XX CC this invention presents a far greater diversity of probes for measuring
XX CC gene expression, with far less bias than expressed sequence tag
XX CC microarrays. The method is suitable for rapid production of functional
XX CC information from genomic sequence. The present sequence is a peptide
XX CC encoded by a single exon nucleic acid probe of the invention. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 76 AA;
XX
Query Match 53.9%; Score 391; DB 4; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.8e-38;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 41 NYVSDGQLSDSSLLFVDPGPMKNVFEPAOKRGICTVHREADFFWISCTADMSLLBQSHS 100
DB 1 NYVSDGQLSDSSLLFVDPGPMKNVFEPAOKRGICTVHREADFFWISCTADMSLLBQSHS 60
QY 101 SPSLYLOCLSKLRQE 116
DB 61 SPSLYLOCLSKLRQE 76
RESULT 5

```

```

AB19775
ID ABB19775 standard; protein; 76 AA.
XX AC ABB19775;
XX DT 23-JAN-2002 (first entry)
XX DE Protein #1774 encoded by probe for measuring heart cell gene expression.
XX KM Human; gene expression; heart; microarray; vascular system;
XX KM cardiovascular disease; hypertension; cardiac arrhythmia;
XX KM congenital heart disease.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488999/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 15; SEQ ID NO 21545; 530bp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart (see
XX CC ABA155-BA41305). The present sequence is a protein encoded by one such
XX CC probe. The probes may be used for predicting, measuring and displaying
XX CC gene expression in samples derived from the human heart via microarrays.
XX CC By measuring gene expression, the probes are useful for predicting,
XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the
XX CC human heart and vascular system e.g. cardiovascular disease,
XX CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 76 AA;
XX
Query Match 53.9%; Score 391; DB 4; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.8e-38;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 41 NYVSDGQLSDSSLLFVDPGPMKNVFEPAOKRGICTVHREADFFWISCTADMSLLBQSHS 100
DB 1 NYVSDGQLSDSSLLFVDPGPMKNVFEPAOKRGICTVHREADFFWISCTADMSLLBQSHS 60
QY 101 SPSLYLOCLSKLRQE 116
DB 61 SPSLYLOCLSKLRQE 76
RESULT 6
AA67546
ID AA67546 standard; protein; 76 AA.
XX AC AA67546;

```

```

XX 06-NOV-2001 (first entry)
DT Human bone marrow expressed probe encoded protein SEQ ID NO: 27852.
XX
XX Human bone marrow expressed exon; gene expression analysis; probe;
DE microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000668.
PR
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 27852; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
XX Sequence 76 AA;
SQ
Query Match 53.9%; Score 391; DB 4; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.8e-38;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 41 NVVSDGQLDESSLLEVDGPMKNVEFFKQKRGCTVHREADPFMSLCTADMSLLEQSHS 100
DB 1 NVVSEGLDESSLLEVDGPMKNVEFFKQKRGCTVHREADPFMSLCTADMSLLEQSHS 60
OY 101 SPSLYIQCLSQKLRQE 116
DB 61 SPSLYIQCLSQKLRQE 76
RESULT 7
AAM55153
ID AAM55153 standard; protein; 76 AA.
XX
XX AAM55153;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 27258.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN

```

```

XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000667.
PR
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
XX Example 4; SEQ ID NO 27258; 650bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
XX
XX Sequence 76 AA;
SQ
Query Match 53.9%; Score 391; DB 4; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.8e-38;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 41 NVVSDGQLDESSLLEVDGPMKNVEFFKQKRGCTVHREADPFMSLCTADMSLLEQSHS 100
DB 1 NVVSEGLDESSLLEVDGPMKNVEFFKQKRGCTVHREADPFMSLCTADMSLLEQSHS 60
OY 101 SPSLYIQCLSQKLRQE 116
DB 61 SPSLYIQCLSQKLRQE 76
RESULT 8
ABG49189
ID ABG49189 standard; peptide; 76 AA.
XX
XX ABG49189;
AC
XX
XX 25-FEB-2003 (first entry)
DT
XX
XX Human liver peptide, SEQ ID NO 27837.
DE
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
OS
XX WO200157273-A2.
XX
XX 03-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000664.
PR
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.

```

PA	(MOE-) MOLECULAR DYNAMICS INC.
XX	
P1	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-476286/51.
PT	Novel single exon nucleic acid probe used to measuring gene expression in
PT	a human breast.
XX	
PS	Claim 27; SEQ ID NO 11858; 322pp; English.
XX	
CC	The present invention relates to novel single exon nucleic acid probes
CC	(see A100010-A110067). The present sequence is a peptide encoded by one
CC	such probe. The probes are useful for measuring human gene expression in
CC	a human breast sample, where the probe hybridises at high stringency to a
CC	nucleic acid expressed in the human breast. The probes are useful for
CC	predicting, diagnosing, grading, staging, monitoring and prognostic
CC	diseases of the human breast, particularly those diseases with polygenic
CC	aetiology. The diseases include: breast cancer, disorders of development,
CC	inflammatory diseases of the breast, fibrocystic changes, proliferative
CC	breast disease and non-carcinoma tumours. Note: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	fip.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 76 AA:
Query Match	53.9%; Score 391; DB 4; Length 76;
Best Local Similarity	98.7%; Pred. No.1.8e-38;
Matches 75; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
OY	41 NYYVSDGLEDSSILEVDGPAMKNVEFKAQRGLCTVHREADFWSICTADMSILBQSHS 100
Db	1 NYVSEGLEDDSSILEVDGPAMKNVEFKAQRGLCTVHREADFWFSICTADMSILBQSHS 60
OY	101 SPSLYLQCSQKLRQE 116
Db	61 SPSLYLQCSQKLRQE 76
RESULT 10	
ABG37140	
ID	ABG37140 standard; peptide; 76 AA.
AC	
XX	ABG37140;
XX	
DT	19-AUG-2002 (first entry)
DE	
XX	Human peptide encoded by genome-derived single exon probe SEQ ID 26805.
XX	
KM	Human; single exon probe; asthma; lung cancer; COPD; IID;
KM	chronic obstructive pulmonary disease; interstitial lung disease;
KM	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KM	tuberosus sclerosis; Gaucher's disease; Niemann-pick disease;
KM	Hernansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KM	pulmonary histiocytosis; lymphangioleiomyomatosis; Karagazer syndrome;
KM	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KM	primary ciliary dyskinesia; pulmonary hypertension;
XX	hyaline membrane disease.
XX	
OS	Homo sapiens.
XX	
PN	WO200186003-A2.
PD	
XX	15-NOV-2001.
PF	
XX	30-JAN-2001; 2001WO-US000665.
PR	
XX	04-FEB-2000; 2000US-0180312P.
ER	26-MAY-2000; 2000US-0207456P.
FR	30-JUN-2000; 2000US-00608408.
RR	03-AUG-2000; 2000US-00632366.
RR	21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
PS Claim 27; SEQ ID NO 26805; 634bp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudaruk syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Kartagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesias, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 76 AA:  
SQ  
Query Match 53.9%; Score 391; DB 5; Length 76;  
Best Local Similarity 98.7%; Pred. No. 1.8e-38;  
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 41 NTVVSGQLFDDSSLLVDDGPAKMGVEFKQKGLCTVHREADFFSLCTAAMSLLEQSHS 100  
DB 1 NTVVSGQLFDDSSLLVDDGPAKMGVEFKQKGLCTVHREADFFSLCTAAMSLLEQSHS 60  
QY 101 SPSLYIQCLSQKXROE 116  
DB 61 SPSLYIQCLSQKXROE 76  
RESULT 11  
AAB14254  
ID AAB14254 standard; protein; 39 AA.  
XX  
AC AAB14254;

XX  
DT 09-FEB-2001 (first entry)  
XX  
XX Partial mouse nedd-2 protein #1.  
DE  
XX  
XX ced-3; virally induced cell death; apoptosis; gene therapy; neural;  
XX muscular degenerative disease; myocardial infarction; stroke; aging;  
XX interleukin-1beta converting enzyme; ICE; mouse; Ice-ced 3 homologue;  
XX Ich; nedd-2.  
XX  
XX Mus sp.  
OS  
XX  
XX Key Location/Qualifiers  
FH 28..32  
FT Domain /label= Active\_domain  
FT  
XX  
XX US6083735-A.  
XX  
XX 04-JUL-2000.  
XX  
XX 10-JUN-1994; 94US-00258287.  
XX  
XX 24-JUN-1993; 93US-00080850.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Yuan J, Miura M;  
XX  
XX WPI; 2000-464343/40.  
XX  
XX N-PSDB; AAA72837.  
XX  
XX New human Ich-1L and Ich-1S proteins for negative and positive regulation  
PT of programmed cell death and for developing therapeutic methods for  
PT diseases and conditions characterized by cell death, e.g. myocardial  
PT infarction or stroke.  
XX  
XX Disclosure; Fig 9; 121bp; English.  
XX  
XX The present sequence is a partial mouse nedd-2 protein. Three possible  
CC reading frames were deduced for the coding sequence of the present  
CC protein. The first reading frame encodes the present protein, which has a  
CC potential QACRG active domain. Nedd-2 is a member of a family of genes  
CC involved in programmed cell death (apoptosis). Other family members  
CC include: the ced-3 gene of C. elegans (AAA72802), human interleukin-1beta  
CC converting enzyme (ICE) (AAB14250), murine ICE1 (AAB14249), human Ich-1  
CC and murine ICE2 (AAB14252). Ich-1 may play an important role in both the  
CC positive and negative regulation of apoptosis. The Ich gene may be used  
CC in gene therapy in disorders characterized by cell death e.g. neural and  
CC muscular degenerative diseases, myocardial infarction, stroke, virally  
CC induced cell death and aging  
XX  
XX Sequence 39 AA:  
SQ  
Query Match 10.1%; Score 73; DB 3; Length 39;  
Best Local Similarity 53.8%; Pred. No. 0.5e;  
Matches 14; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
QY 15 LHHIRRMFGDSCPYLACKPMPFFIQ 40  
DB 3 LOEVRRLFDNANCPGLQNKPPMPFFIQ 28  
RESULT 12  
ADE08128  
ID ADE08128 standard; protein; 101 AA.  
XX  
XX ADE08128;  
XX  
XX 29-JUN-2004 (first entry)  
DT  
XX  
XX Novel protein (useful for identifying genetic disorders) #283.  
DE  
XX novel gene; novel protein; tissue marker; molecular weight marker;  
KM

XX chromosome marker; genetic disorder.

XX Unidentified.

XX MO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002MO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-037281P.

XX 12-APR-2002; 2002US-037281P.

XX 22-APR-2002; 2002US-0012855P.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RM, Ren F, Zhang J, Zhao QR, Wang J,

XX Gosh W, Xue AJ, Wehrman T, Wang G, Zhou P, Drmanac RT, Wang Z;

XX Ma Y, Wang D, Chen R, Xu C, Boyle BJ,

XX WPI; 2003-569235/53.

XX N-PSDB; ADS07217.

XX New polynucleotides, useful for expressing recombinant proteins for

XX PT analysis, characterization or therapeutic use, or as markers for tissues

XX in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1194; 1177bp; English.

XX The invention comprises the amino acid and coding sequences of novel

XX CC proteins. The DNA and protein sequences of the invention are useful as:

XX CC markers for tissues in which the corresponding protein is preferentially

XX CC expressed; as molecular weight markers on gels; as chromosome markers or

XX CC tags; to identify chromosomes or to map related gene positions; and to

XX CC compare with endogenous DNA sequences in patients to identify potential

XX CC genetic disorders. The present amino acid sequence represents a protein

XX CC of the invention.

XX SQ Sequence 101 AA;

XX Query Match 8.5%; Score 61.5; DB 7; Length 101;

XX Best Local Similarity 26.0%; Pred. No. 45;

XX Matches 26; Conservative 14; Mismatches 27; Indels 33; Gaps 5;

XX QY 12 GLPLHIRMFMGDSCPYLAKRPMFPIQNYVVSQGLSDSLLEVDGP--AMKNVEFK 68

XX Db 18 GLPLHIRMFMGDSCPYLAKRPMFPIQNYVVSQGLSDSLLEVDGP--AMKNVEFK 68

XX QY 69 AQRKGLCTVHREADPFWMSICTADMSILQSHSPSLVYQC 108

XX Db 57 GTRK-----RNRQDAW--IADYLCPCQTUTPLWLRKC 87

XX RESULT 13

XX ID AAU14746 standard; protein; 80 AA.

OS Homo sapiens.

XX MO200157187-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001MO-US003782.

XX 03-FEB-2000; 2000US-00496914.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 30-NOV-2000; 2000US-0250683P.

XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ,

XX Ren F, Drmanac RT;

XX WPI; 2001-488875/53.

XX N-PSDB; AAS23051.

XX Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and

XX PT gene therapy.

XX Claim 10; Page 135; 392pp; English.

XX AAU14602-AAU14794 represent novel bone marrow polypeptides of the

XX CC invention. The proteins and corresponding coding sequences may be used in

XX CC the prevention, diagnosis and treatment of diseases associated with

XX CC inappropriate bone marrow polypeptide expression. For example, to treat

XX CC disorders associated with decreased expression by rectifying mutations or

XX CC deletions in a patient's genome that affect the activity of the

XX CC polypeptides by expressing inactive proteins or to supplement the

XX CC patient's own production of the polypeptide. Additionally, the nucleic

XX CC acids may be used to produce the polypeptides, by inserting the nucleic

XX CC acids into a host cell and culturing the cell to express the protein. The

XX CC nucleic acid and its complementary sequences may also be used as DNA

XX CC probes in diagnostic assays to detect and quantitate the presence of

XX CC similar nucleic acid sequences in samples, and therefore which patients

XX CC may be in need of restorative therapy. The proteins may also be used as

XX CC antigens in the production of antibodies against bone marrow proteins and

XX CC in assays to identify modulators of their expression and activity. The

XX CC anti-bone marrow protein antibodies and antagonists may also be used to

XX CC down regulate expression and activity. The antibodies may also be used as

XX CC diagnostic agents for detecting the presence of the protein in samples

XX CC (e.g. by enzyme linked immunosorbent assay (ELISA)). The proteins may be

XX CC used to regulate haematopoiesis activity, and consequently in the

XX CC treatment of myeloid or lymph cell disorders; in tissue regeneration,

XX CC such as wound healing; as a nutritional supplement; and in treatment of

XX CC immune disorders such as severe combined immunodeficiency (SCID)

XX SQ Sequence 80 AA;

XX Query Match 8.3%; Score 60.5; DB 4; Length 80;

XX Best Local Similarity 32.8%; Pred. No. 43;

XX Matches 20; Conservative 7; Mismatches 19; Indels 15; Gaps 2;

XX QY 17 HIRMFMDSCPYLAKRPMFPIQNYVVSQGLSDSLLEVDGPAMKNVEFKRGLCT 76

XX Db 23 HSRRLKTVDSAPSTAGD-----VMADLQWDS-----GNGENLVAVGAERGRHT 67

XX QY 77 V 77

XX Db 68 M 68

XX RESULT 14

XX ID AAU40343 standard; protein; 72 AA.

XX AC AAU40343;

XX DT 13-FEB-2002 (first entry)

```

XX DE Propionibacterium acnes immunogenic protein #1239.
XX KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KM dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001MO-US012865.
XX PR 21-APR-2000; 2000US-0199047P.
XX PR 02-JUN-2000; 2000US-020841P.
XX PR 07-JUL-2000; 2000US-0216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skelky YW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L malsommeuve J, Zhang Y, Jen S, Carter D;
XX DR WPI; 2001-616774/71.
XX DR N-PSDB; AAS59511.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris.
XX PS Example 1; SEQ ID NO 1538; 1069pp; English.
XX XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
XX CC this patent did not form part of the printed specification, but was
XX CC obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 72 AA;
XX
XX Query Match 8.3%; Score 60; DB 4; Length 72;
XX Best Local Similarity 35.4%; Pred. No. 43;
XX Matches 17; Conservative 4; Mismatches 23; Indels 4; Gaps 2;
XX
XX QY 2 SVYGVQDTHSGPLPHIRMFMDSCPVLAKGPKMFIONVYVSDGOL 49
XX Db 14 STYATATATKAG-PIRIRITRLTSPSCPLRGAPKR---HNCHPSGGEL 57
XX
XX RESULT 15
XX ABM36862
XX ID ABM36862 standard; protein; 72 AA.
XX AC ABM36862;
XX XX
XX DT 20-OCT-2003 (first entry)

```

```

XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #1538.
XX KM Acne vulgaris; antiseborrheic; dermatological; antibacterial;
XX KM immunostimulant; immune response; vaccine.
XX OS Propionibacterium acnes.
XX PN WO2003033515-A1.
XX PD 24-APR-2003.
XX PF 11-OCT-2002; 2002MO-US0322727.
XX PR 15-OCT-2001; 2001US-00978825.
XX PA (CORI-) CORIXA CORP.
XX PI Mitcham JL, Skelky YW, Persing DH, Bhatia A, Malsommeuve JL;
XX PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
XX PI Barth B, Vallieve-Douglas J;
XX DR WPI; 2003-381789/36.
XX DR N-PSDB; ACP64440.
XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
XX PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
XX PT or for stimulating an immune response specific for a P. acnes protein.
XX PS Example 1; SEQ ID NO 1538; 1481pp; English.
XX XX The invention relates to an isolated polynucleotide (ACP64435-ACP64733)
XX CC encoding a Propionibacterium acnes protein. The invention also relates to
XX CC polypeptides encoded by the polynucleotides (ABM55624-ABM64536) and to
XX CC immunogenic fragments of P. acnes polypeptides. The invention
XX CC additionally encompasses expression vectors and host cells comprising a
XX CC polynucleotide of the invention; antibodies against polypeptides of the
XX CC invention; fusion proteins comprising a polypeptide of the invention; a
XX CC method for stimulating an immune response specific for a P. acnes
XX CC polypeptide and an isolated T cell population comprising P. acnes
XX CC prepared via this method; a vaccine composition (comprising P. acnes polypeptides,
XX CC polynucleotides, antibodies, fusion proteins, T cell populations, or
XX CC antigen-presenting cells that express the polypeptide); a method and kit
XX CC for detecting or determining the presence or absence of P. acnes in a
XX CC patient; and a method for inhibiting the development of P. acnes in a
XX CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
XX CC proteins, T cell populations or antigen-presenting cells that express the
XX CC polypeptides are useful for diagnosing, preventing or treating acne
XX CC vulgaris, or for stimulating an immune response specific for a P. acnes
XX CC protein. The polynucleotides can also be used as probes or primers for
XX CC nucleic acid hybridization. The vaccine composition is useful for the
XX CC stimulation of an immune response against P. acnes, or for treating acne,
XX CC and the kit is useful for performing a diagnostic assay. The present
XX CC sequence represents a polypeptide predicted to be encoded by an ORF (open
XX CC reading frame) contained within the P. acnes polynucleotides of the
XX CC invention. Note: The sequence data for this patent did not form part of
XX CC the printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 72 AA;
XX
XX Query Match 8.3%; Score 60; DB 6; Length 72;
XX Best Local Similarity 35.4%; Pred. No. 43;
XX Matches 17; Conservative 4; Mismatches 23; Indels 4; Gaps 2;
XX
XX QY 2 SVYGVQDTHSGPLPHIRMFMDSCPVLAKGPKMFIONVYVSDGOL 49
XX Db 14 STYATATATKAG-PIRIRITRLTSPSCPLRGAPKR---HNCHPSGGEL 57
XX
XX Search completed: September 9, 2004, 06:24:41
XX Job time : 56 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 06:23:44 ; Search time 46 Seconds

(without alignments)  
955.097 Million cell updates/sec

Title: US-10-713-208-6\_COPY\_205\_341

Sequence: 1 QSVGVGDTHSGPLDHLHRR.....GTPGSGITSCKMHPSSIG 137

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 655639

Minimum DB seq length: 0

Maximum DB seq length: 137

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	53.9	76	US-09-864-761-35073	Sequence 35073, A
2	77.5	10.7	56	US-09-989-903-38	Sequence 58, Appl
3	77.5	10.7	56	US-10-068-564-58	Sequence 58, Appl
4	77	10.6	52	US-09-989-903-51	Sequence 51, Appl
5	77	10.6	52	US-10-068-564-51	Sequence 51, Appl
6	73	10.1	39	US-08-459-455-45	Sequence 45, Appl
7	70.5	9.7	51	US-09-989-903-45	Sequence 45, Appl
8	70.5	9.7	51	US-10-068-564-45	Sequence 45, Appl
9	68	9.4	39	US-09-989-903-38	Sequence 38, Appl
10	68	9.4	39	US-10-068-564-38	Sequence 38, Appl
11	66	9.1	39	US-09-989-903-30	Sequence 30, Appl
12	66	9.1	39	US-10-068-564-30	Sequence 30, Appl
13	64	8.8	95	US-10-767-701-58487	Sequence 58487, A
14	62	8.6	94	US-10-424-599-187873	Sequence 187873, A
15	62	8.6	94	US-10-424-599-193827	Sequence 193827, A

16	61	8.4	95	US-10-424-599-192160	Sequence 192160, A
17	61	8.4	131	US-10-437-963-134692	Sequence 134692, A
18	60.5	8.3	86	US-10-437-963-140300	Sequence 140300, A
19	60.5	8.3	95	US-10-424-599-204386	Sequence 204386, A
20	60	8.3	85	US-10-424-599-207479	Sequence 207479, A
21	59.5	8.2	131	US-09-839-185-6	Sequence 6, Appl
22	59.5	8.2	131	US-10-412-699B-302	Sequence 302, Appl
23	59	8.1	111	US-10-767-701-42670	Sequence 42670, A
24	58.5	8.1	67	US-10-437-963-128338	Sequence 128338, A
25	58.5	8.1	106	US-10-424-599-201058	Sequence 201058, A
26	58	8.0	108	US-10-424-599-178808	Sequence 178808, A
27	58	8.0	108	US-10-424-599-247725	Sequence 247725, A
28	57.5	7.9	87	US-10-437-963-191300	Sequence 191300, A
29	57.5	7.9	86	US-10-437-963-191300	Sequence 191300, A
30	57	7.9	104	US-10-424-599-208754	Sequence 208754, A
31	56.5	7.8	110	US-10-424-599-245890	Sequence 245890, A
32	56	7.7	106	US-10-437-963-131533	Sequence 131533, A
33	56	7.7	106	US-10-437-963-162845	Sequence 162845, A
34	56	7.7	106	US-10-437-963-108763	Sequence 108763, A
35	56	7.7	131	US-10-767-701-62785	Sequence 62785, A
36	55.5	7.7	97	US-10-437-963-126887	Sequence 126887, A
37	55.5	7.7	97	US-10-767-701-39413	Sequence 39413, A
38	55.5	7.7	114	US-10-424-599-224073	Sequence 224073, A
39	55.5	7.7	129	US-10-767-701-32508	Sequence 32508, A
40	55	7.6	133	US-10-393-840-45	Sequence 45, Appl
41	55	7.6	78	US-10-437-963-102975	Sequence 102975, A
42	55	7.6	103	US-10-437-963-149012	Sequence 149012, A
43	55	7.6	113	US-10-282-122A-57447	Sequence 57447, A
44	54.5	7.5	89	US-10-767-701-34958	Sequence 34958, A
45	54.5	7.5	107	US-10-424-599-194579	Sequence 194579, A
			12	US-10-425-114-36601	Sequence 36601, A

## ALIGNMENTS

RESULT 1  
US-09-864-761-35073  
; Sequence 35073, Application US/09864761  
; Patent No. US00020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wenheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBS USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35073
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007283.3
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.9
; OTHER INFORMATION: EST_HUMAN HIT: A1139524.1, EVALU2.00e-38
US-09-864-761-35073

```

```

Query Match      53.9%; Score 391; DB 9; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.8e-38;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      41 NYVSDGQLEDSLLLEVDGPAMKNVFEKQKQGLCTVHREADFFNSLCTADMSLLEQSHS 100
DB      1 NYVSEGLDEDSLLLEVDGPAMKNVFEKQKGLCTVHREADFFNSLCTADMSLLEQSHS 60
QY      101 SPSYIQLCSQKLRQE 116
DB      61 SPSLYIQLCSQKLRQE 76

```

```

RESULT 2
US-09-989-903-58
; Sequence 58, Application US/09989903
; Patent No. US20020146804A1
; GENERAL INFORMATION:
; APPLICANT: Alnemati, Emed S.
; APPLICANT: Fernandez-Alnemati, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434D1
; CURRENT APPLICATION NUMBER: US/09/989,903
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-989-903-58

```

```

Query Match      10.7%; Score 77.5; DB 9; Length 56;
Best Local Similarity 40.4%; Pred. No. 0.22;
Matches 19; Conservative 5; Mismatches 20; Indels 3; Gaps 1;
QY      13 LPLHHRFMGDSCEPYLAGKPKMFQIYVYVSDGLEDSLLEVDG 59
DB      2 LQLQEVFRFLFDNANCPSLQNKPKKFFIQ--ACRGDETDRGVDDQG 45

```

```

RESULT 3
US-10-068-564-58
; Sequence 58, Application US/10068564
; Publication No. US20030040096A1
; GENERAL INFORMATION:
; APPLICANT: Alnemati, Emed S.
; APPLICANT: Fernandez-Alnemati, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434C2
; CURRENT APPLICATION NUMBER: US/10/068,564
; CURRENT FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-068-564-58

```

```

Query Match      10.7%; Score 77.5; DB 14; Length 56;
Best Local Similarity 40.4%; Pred. No. 0.22;
Matches 19; Conservative 5; Mismatches 20; Indels 3; Gaps 1;

```

```

QY      13 LPLHHRFMGDSCEPYLAGKPKMFQIYVYVSDGLEDSLLEVDG 59
DB      2 LQLQEVFRFLFDNANCPSLQNKPKKFFIQ--ACRGDETDRGVDDQG 45

```

```

RESULT 4
US-09-989-903-51
; Sequence 51, Application US/09989903
; Patent No. US20020146804A1
; GENERAL INFORMATION:
; APPLICANT: Alnemati, Emed S.
; APPLICANT: Fernandez-Alnemati, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434D1
; CURRENT APPLICATION NUMBER: US/09/989,903
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-989-903-51

```

```

Query Match      10.6%; Score 77; DB 9; Length 52;
Best Local Similarity 50.0%; Pred. No. 0.23;
Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
QY      15 LHHRRFMGDSCEPYLAGKPKMFQIYVYVSDGLEDSLLEVDG 40
DB      4 IYDLTSYFTGSKCPSLSGKPKIFIQ 29

```

```

RESULT 5
US-10-068-564-51
; Sequence 51, Application US/10068564
; Publication No. US20030040096A1
; GENERAL INFORMATION:
; APPLICANT: Alnemati, Emed S.
; APPLICANT: Fernandez-Alnemati, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434C2
; CURRENT APPLICATION NUMBER: US/10/068,564
; CURRENT FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 78

```

SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 51  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-068-564-51

Query Match 10.6%; Score 77; DB 14; Length 52;  
Best Local Similarity 50.0%; Pred. No. 0.23;  
Matches 13; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 15 LHHIRFMGDSCEPYLAKRPMFFIQ 40  
DB 4 IYDHSYFTGSKCPSLGSKPKIFIIQ 29

RESULT 6  
US-08-459-455-45  
Sequence 45, Application US/08459455  
Publication No. US20030124105A1  
GENERAL INFORMATION:  
APPLICANT: Yuan, Junying  
APPLICANT: Mura, Masayuki  
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins  
NUMBER OF SEQUENCES: 95  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,455  
FILING DATE: 2-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/368,704  
FILING DATE: 4-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/258,287  
FILING DATE: 10-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/080,850  
FILING DATE: 24-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Bugalsky, Lawrence B.  
REGISTRATION NUMBER: 35,086  
REFERENCE/DOCKET NUMBER: 0609.3920003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-459-455-45

Query Match 10.1%; Score 73; DB 8; Length 39;  
Best Local Similarity 53.8%; Pred. No. 0.47;  
Matches 14; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 15 LHHIRFMGDSCEPYLAKRPMFFIQ 40

DB 3 LQEVFRFLPDNANCSLQNRKPMFFIQ 28

RESULT 7  
US-09-989-903-45  
Sequence 45, Application US/09989903  
Patent No. US20020146804A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emed S.  
APPLICANT: Fernandez-Alnemri, Teresa  
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
FILE REFERENCE: 480140.434D1  
CURRENT APPLICATION NUMBER: US/09/989,903  
CURRENT FILING DATE: 2002-04-11  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 45  
LENGTH: 51  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-989-903-45

Query Match 9.7%; Score 70.5; DB 9; Length 51;  
Best Local Similarity 37.3%; Pred. No. 1.3;  
Matches 19; Conservative 5; Mismatches 16; Indels 11; Gaps 2;

QY 8 QHSGLEPLHHRMFMGDSCEPYLAKRPMFFIYVYVSDGLEDSLLFVD 58  
DB 4 QTLTGL-----FKGDKCSLVGKPKIFIIQ---ACRGSQHDVPVPLD 43

RESULT 8  
US-10-068-564-45  
Sequence 45, Application US/10068564  
Publication No. US20030040096A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emed S.  
APPLICANT: Fernandez-Alnemri, Teresa  
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
FILE REFERENCE: 480140.434C2  
CURRENT APPLICATION NUMBER: US/10/068,564  
CURRENT FILING DATE: 2002-02-05  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 45  
LENGTH: 51  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-068-564-45

Query Match 9.7%; Score 70.5; DB 14; Length 51;  
Best Local Similarity 37.3%; Pred. No. 1.3;  
Matches 19; Conservative 5; Mismatches 16; Indels 11; Gaps 2;

QY 8 QHSGLEPLHHRMFMGDSCEPYLAKRPMFFIYVYVSDGLEDSLLFVD 58  
DB 4 QTLTGL-----FKGDKCSLVGKPKIFIIQ---ACRGSQHDVPVPLD 43

RESULT 9  
US-09-989-903-38  
Sequence 38, Application US/09989903  
Patent No. US20020146804A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emed S.  
APPLICANT: Fernandez-Alnemri, Teresa  
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING  
FILE REFERENCE: 480140.434D1  
CURRENT APPLICATION NUMBER: US/09/989,903

/ CURRENT FILING DATE: 2002-04-11  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: Patentn Ver. 2.0  
/ SEQ ID NO: 38  
/ LENGTH: 39  
/ TYPE: PRT  
/ ORGANISM: Mus musculus  
US-09-989-903-38

Query Match 9.4%; Score 68; DB 9; Length 39;  
Best Local Similarity 48.1%; Pred. No. 1.8;  
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 14 PLHIRMFMGDSCPYLAGKPKMFFIQ 40  
Db 2 PIKDLTAHFGRDCKTLEKPKLFFIQ 28

RESULT 10  
US-10-068-564-38  
/ Sequence 38, Application US/10068564  
/ Publication No. US20030040096A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Alnemri, Emed S.  
/ TITLE OF INVENTION: CASPASE-14, AN APOTOTIC PROTEASE, NUCLEIC ACID ENCODING  
/ FILE REFERENCE: 480140.434C2  
/ CURRENT FILING DATE: 2002-02-05  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: Patentn Ver. 2.0  
/ SEQ ID NO: 38  
/ LENGTH: 39  
/ TYPE: PRT  
/ ORGANISM: Mus musculus  
US-10-068-564-38

Query Match 9.4%; Score 68; DB 14; Length 39;  
Best Local Similarity 48.1%; Pred. No. 1.8;  
Matches 13; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 14 PLHIRMFMGDSCPYLAGKPKMFFIQ 40  
Db 2 PIKDLTAHFGRDCKTLEKPKLFFIQ 28

RESULT 11  
US-09-989-903-30  
/ Sequence 30, Application US/09989903  
/ Patent No. US20020146804A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Alnemri, Emed S.  
/ APPLICANT: Fernandez-Alnemri, Teresa  
/ TITLE OF INVENTION: CASPASE-14, AN APOTOTIC PROTEASE, NUCLEIC ACID ENCODING  
/ FILE REFERENCE: 480140.434D1  
/ CURRENT FILING DATE: 2002-04-11  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: Patentn Ver. 2.0  
/ SEQ ID NO: 30  
/ LENGTH: 39  
/ TYPE: PRT  
/ ORGANISM: Mus musculus  
US-09-989-903-30

Query Match 9.1%; Score 66; DB 9; Length 39;  
Best Local Similarity 50.0%; Pred. No. 3.2;  
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 15 LHHIRMFMGDSCPYLAGKPKMFFIQ 40

Db 3 LKXLTSPFRGDYCRSLTGKPKLFFIQ 28

RESULT 12  
US-10-068-564-30  
/ Sequence 30, Application US/10068564  
/ Publication No. US20030040096A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Alnemri, Emed S.  
/ APPLICANT: Fernandez-Alnemri, Teresa  
/ TITLE OF INVENTION: CASPASE-14, AN APOTOTIC PROTEASE, NUCLEIC ACID ENCODING  
/ FILE REFERENCE: 480140.434C2  
/ CURRENT FILING DATE: 2002-02-05  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: Patentn Ver. 2.0  
/ SEQ ID NO: 30  
/ LENGTH: 39  
/ TYPE: PRT  
/ ORGANISM: Mus musculus  
US-10-068-564-30

Query Match 9.1%; Score 66; DB 14; Length 39;  
Best Local Similarity 50.0%; Pred. No. 3.2;  
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 15 LHHIRMFMGDSCPYLAGKPKMFFIQ 40  
Db 3 LKXLTSPFRGDYCRSLTGKPKLFFIQ 28

RESULT 13  
US-10-767-701-58487  
/ Sequence 58487, Application US/10767701  
/ Publication No. US20040172684A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Kovalic, David K.  
/ APPLICANT: Zhou, Yihua  
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
/ FILE REFERENCE: 38-21(5353) B  
/ CURRENT FILING DATE: 2004-01-29  
/ NUMBER OF SEQ ID NOS: 63128  
/ SEQ ID NO: 58487  
/ LENGTH: 95  
/ TYPE: PRT  
/ ORGANISM: Sorghum bicolor  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: 30980100.pep  
US-10-767-701-58487

Query Match 8.8%; Score 64; DB 16; Length 95;  
Best Local Similarity 27.6%; Pred. No. 18;  
Matches 24; Conservative 14; Mismatches 33; Indels 16; Gaps 4;

Qy 6 VQHTSGLEPLHIRMFMGDS-----CPYLAGKPKMFF-----IQTYVSDGQLEDS 52  
Db 1 INQTKSLQLVKVRHQAATYASDFPLLEKFORFTRSLKTVSESLVGSNDALDS 60

Qy 53 SLEVDGPAMKQVEFKAKGGLC-TVH 78  
Db 61 A--SDSPSVKMKMKSKVACLTIX 85

RESULT 14  
US-10-424-599-187873  
/ Sequence 187873, Application US/10424599  
/ Publication No. US20040031072A1  
/ GENERAL INFORMATION:  
/ APPLICANT: La Rosa Thomas J

```
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 187873
/ LENGTH: 94
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_140662C.1.pep
US-10-424-599-187873
```

Query Match 8.6%; Score 62; DB 12; Length 94;

Best Local Similarity 25.3%; Pred. No. 31; Mismatches 31; Indels 24; Gaps 5;

```
QY 40 QNYVSDGLEDSSLEVDGPAKQVFEKAKRGCTVHREADPFW-----SLCTADMS 93
DB 1 QDPVID-----MURVDVSMVNL-----KREHSILRKCDVVFPISDLSLC-----S 45
QY 94 LLEQSHSPSLY--QCISQTLROERGTFPGSGITBSK 129
DB 46 LQMEVRSscyIHSDCIDRVLRSSSSLRMATTTGAR 84
```

## RESULT 15

```
US-10-424-599-193827
/ Sequence 193827, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 193827
/ LENGTH: 94
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_17050C.1.pep
US-10-424-599-193827
```

Query Match 8.6%; Score 62; DB 12; Length 94;

Best Local Similarity 21.8%; Pred. No. 31; Mismatches 26; Indels 16; Gaps 2;

```
QY 45 SDGLEDSSLEVD-----GPAKQVFEKAKRGCTVHRE-----ADPFWNSLC 88
DB 15 SNGEIVKSGVALEKIVATERYAGTSWHELVITRQALGFLVHOKRNXSLERIDLC 74
QY 89 TADMSLLEQSHSPSLY 106
DB 75 PAVISLIPSVHSPPTATYM 92
```

Search completed: September 9, 2004, 06:27:01  
Job time : 47 secs

*This Page Blank (uspto)*